bioRxiv preprint doi: https://doi.org/10.1101/2020.08.19.258095; this version posted August 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

BART3D: Inferring transcriptional regulators associated with differential chromatin interactions from Hi-C data

Zhenjia Wang[1], Yifan Zhang[1,2], Chongzhi Zang[1,2,3]*

[1] Center for Public Health Genomics, [2] Department of Biomedical Engineering, [3] Department of Public Health Sciences, and Department of Biochemistry and Molecular Biology, University of Virginia, Charlottesville, VA 22908, USA

* To whom correspondence should be addressed: zang@virginia.edu

ABSTRACT

Identification of functional transcriptional regulators (TRs) associated with chromatin interactions is an important problem in studies of 3-dimensional genome organization and gene regulation. Direct inference of TR binding has been limited by the resolution of Hi-C data. Here, we present BART3D, a computational method for inferring TRs associated with genome-wide differential chromatin interactions by comparing two Hi-C maps, leveraging public ChIP-seq data for human and mouse. We demonstrate that BART3D can detect target TRs inducing chromatin architecture changes from dynamic Hi-C profiles with TR perturbation. BART3D can be a useful tool in 3D genome data analysis and functional genomics research.

KEYWORDS

Hi-C, transcriptional regulator, ChIP-seq, chromatin architecture, protein-DNA interaction

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.19.258095; this version posted August 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

INTRODUCTION

The three-dimensional (3D) organization of eukaryotic genomes affects transcriptional gene regulation [1-3]. Although topologically associating domains (TADs) appear to be conserved across cell types at the level of cell populations [4-6], chromatin architecture is highly dynamic during development and cell differentiation [7,8], and can be disrupted in disease states [1]. Transcriptional regulators (TRs), including transcription factors and chromatin regulators, are required for the establishment and maintenance of chromosomal architecture [9-11]. Identification of functional TRs associated with chromatin dynamics can help unravel the spatial organization of the genome and the impact of 3D architecture on gene regulation. The 3D organization of the genome can be measured using chromosome conformation capture-based methods such as Hi-C [12] and in situ Hi-C [5]. Chromatin interaction events can be detected by inferring loop structures from signal enrichment in Hi-C contact maps [5,13]. However, limited by the restriction enzyme digestion and ligation procedure in experiments and highly dependent on the sequencing depth, the resolution of Hi-C maps is typically 10^4-10^5 bp, or can be as high as 10^3 bp for ultra-deep *in situ* Hi-C [5]. It is still difficult to reach the sub-nucleosomal resolution of TR binding events (10^1-10^2 bp). HiChIP [14] and PLAC-seq [15] can reach higher resolution for easier association with TR binding but require additional experimental steps to use a preselected protein factor as an anchor, limiting the feasibility for an unbiased TR association analysis. A computational method to identify TR binding directly from lower-resolution Hi-C data is necessary for functional analysis of 3D genome data.

Most computational methods for differential Hi-C data analysis, including diffHic [16], FIND [17], HiCcompare [18], and Selfish [19], focus on detecting changes in chromatin interaction events on the locus-to-locus level by comparing Hi-C data generated in two biological states. Few methods aim to generate a profile of differential interaction across the whole genome to associate with TR binding profiles. TR inference from collected genomic binding profiles is a

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.19.258095; this version posted August 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

more powerful approach than conventional DNA sequence motif search [20-22]. We previously developed BART [20], an algorithm for inferring TRs whose binding profiles associate with a query genomic profile, leveraging over 13,000 human and mouse ChIP-seq datasets collected from the public domain. Here, we present BART3D, a new bioinformatics tool for 3D genome data analysis built upon the BART algorithm. BART3D integrates Hi-C maps with public ChIP-seq data to infer TRs associated with genome-wide changes in chromatin interactions.

RESULTS

Overall design of BART3D

Starting with two Hi-C-type 3D genome contact maps as input, BART3D first generates a genomic differential chromatin interaction (DCI) profile by comparing the two contact maps, then uses the BART algorithm [20] to identify transcriptional regulators (TRs) whose binding sites are associated with either increased or decreased chromatin interactions (Fig. 1). BART3D can accept three formats of unnormalized genomic contact maps as input: 1) raw count matrices from HiC-Pro [23], 2) .hic format files from Juicer [24], and 3) .cool format files [25]. The output is a ranked list of TRs with a series of statistical measurements.

We employ an innovative approach to quantify the difference between two Hi-C contact matrices under different conditions, e.g., treatment and control. To account for the negative correlation between intra-chromosomal interaction frequency and genomic distance (Supplementary Fig. S1) [12] and to extract chromatin architecture information, we first normalize the contact matrix of each chromosome using a distance-based approach, where the read count in each bin pair is normalized by the average read count across all bin pairs at the same genomic distance (Supplementary Fig. S2). For any region in a chromosome, we consider the intra-chromosomal interactions between this region and its flanking regions within a certain genomic distance, e.g., 500 kb, quantified by an array of contact scores in a column/row of the

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.19.258095; this version posted August 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

contact matrix, represented by the dashed, blue, 45-degree boxes in Fig. 1a. We assess the DCI for this region by comparing the arrays of contact scores from the two conditions, e.g., treatment and control (Fig. 1a). The DCI score is calculated as the paired t-test statistic between the two sets of contact scores:

$$DCI = \frac{\bar{d}}{s_d/\sqrt{n}}$$

where $\bar{d}$ and s_d are the mean and the standard deviation, respectively, of $\{d_i\}, i = 1, \ldots, n$, defined as the difference in contact scores between treatment and control, calculated by subtracting the paired elements from the two arrays, and *n* is the length of the contact score arrays. Positive or negative DCI scores represent increased or decreased chromatin interactions, respectively, from control to treatment. We generate a genome-wide DCI profile by scanning all chromosomes to calculate DCI scores for every non-overlapping bin across each chromosome (Fig. 1b).

We then infer TRs whose genome-wide binding profiles are associated with the DCI profile derived from Hi-C contact matrices (Fig. 1c,d). We map the genomic DCI profile to the union DNaseI hypersensitive sites (UDHS), a curated dataset representing all putative cis-regulatory elements (CREs) in the genome [26], and generate a cis-regulatory profile in which the score for each candidate CRE is set to equal the DCI score of the genomic region in which the CRE is located. We use the BART algorithm [20] to infer TRs that preferentially bind at CREs with a high score, representing increased chromatin interactions. Then we flip the DCI profile and perform BART analysis again to infer TRs whose binding are associated with decreased chromatin interactions (Fig. 1e).

Performance of distance-based normalization

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.19.258095; this version posted August 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

We first evaluated the feasibility and performance of the distance-based normalization method used in BART3D. By targeting a specific factor of interest, HiChIP signals are enriched at the target-bound loci [14]. With appropriate normalization, the HiChIP target factor should be inferred from a static HiChIP dataset. Using this measure, we collected 84 HiChIP datasets targeting different TRs and tested how different normalization methods affect the inference of the HiChIP target factor. For each HiChIP dataset, we generated genomic contact maps at 5kb resolution without normalization, with ICE normalization [23], and with distance-based normalization. We then generated a genomic profile from each contact map, in which each 5kb bin across the genome is scored as the sum of interaction signals between this bin and all of its flanking bins within 500 kb. We used BART to infer TRs associated with this genomic interaction profile. As a control, we also ran BART analysis on the HiChIP sequence read pile-up profile and used the rank of the target factor as a reference. Overall, the distance-based normalization yielded high ranks of the HiChIP factors comparable to the HiChIP profile control, higher than either unnormalized or ICE normalized approaches (Fig. 2). This result indicates that the distance-based normalization is an optimal approach for quantifying Hi-C matrices for BART3D analysis.

BART3D can identify perturbed TRs from dynamic Hi-C data

To demonstrate the performance of BART3D, we calculated DCI profiles and inferred TRs for several published Hi-C experiments comparing wild type (WT) with DNA-associating factor knockout (KO) models in mouse cells. The KO targets include transcription factor Ctcf [27] and cohesin complex component Rad21 [28], known to function cooperatively to induce DNA looping and maintain TAD structures [29]; as well as Nr1d1 [30], and Smchd1 [31], both of which have repressive effects on transcriptional regulation and chromatin architecture. As expected, genomic regions containing binding sites of CTCF or RAD21 exhibit decreased chromatin interaction levels after KO (Fig. 3a,b), while those containing NR1D1 or SMCHD1

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.19.258095; this version posted August 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

sites associate with increased chromatin interactions after KO in their corresponding samples (Fig. 3c,d). This result indicates that the DCI profile can connect perturbed protein binding sites with differential chromatin interaction. Indeed, the BART3D results show that the KO factors are always among the top-ranked TRs inferred to be associated with the corresponding decreased or increased chromatin interactions for all four cases (Fig. 4, labeled in red). These results show that BART3D can successfully infer TRs that induce chromatin interaction changes from Hi-C data.

In addition to the KO factor itself, we also found other TRs highly ranked in the BART3D results from the KO/WT Hi-C comparisons (Fig. 4). For Ctcf or Rad21 KO (Fig. 4a,b), several top inferred TRs, including SMC1A, SMC3, and STAG1 are all components of the cohesin complex [32]. For Nr1d1 KO (Fig. 4c), other top-ranked TRs are either co-regulators (PPARA, HNF4A, and FOXA2) or a target (ARNTL) of NR1D1, which regulates metabolic and circadian pathways [33-36]. For Smchd1 KO (Fig. 4d), SUZ12, EZH2, BMI1, and RNF2 are all related to polycomb group (PcG) factors, which have been shown to interact with SMCHD1 [37] and have repressive effects on transcription and chromatin state. These results indicate that BART3D can help identify novel TRs with biological relevance by mining Hi-C data.

DISCUSSION

We developed BART3D for differential analysis of Hi-C data and to infer functional TRs associated with changes in chromatin interactions. BART3D overcomes the relatively low resolution of Hi-C data and connects chromatin interactions on the multi-kb to Mb level to cis-regulatory events on the nucleosomal or base-pair level by accounting for statistical differences in Hi-C signals within a large distance range and using a predefined genomic CRE set. BART3D uses a distance-based normalization approach, which can remove cross-sample biases (Supplementary Fig. S2) and outperforms ICE normalization [23] in detecting local chromatin

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.19.258095; this version posted August 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

interactions (Fig. 2). We use dynamic Hi-C datasets from TR KO experiments to show that BART3D can accurately infer TRs inducing chromatin architecture changes as well as other TRs with biological relevance.

In the framework of BART3D, we assume that differential chromatin interaction is mainly caused by changes in genomic binding of transcriptional regulator proteins, which act primarily in cis. Other events that can also result in pattern changes on Hi-C maps such as structural variations in the genome are not considered in BART3D. Under this assumption, we focus on intra-chromosomal interactions within a certain range of chromosomal distance and completely ignore inter-chromosomal interactions. The default genomic distance is set as 200 kb, but users can adjust this parameter in exploratory studies for optimizing discovery power, as different TRs may associate with chromatin interactions at different genomic ranges (Fig. 5). There are not many other tunable parameters. The bin size should be consistent with the Hi-C contact maps under interrogation and is restricted to the Hi-C data resolution. Replicates of Hi-C data are not accounted for in the DCI calculation but can be used to generate a background control for TR inference, i.e., TRs inferred from comparing replicates of Hi-C data from the same biological condition are likely due to technical variations and can be considered as false positive. Such TRs should be discarded if they also appear in results from cross-condition Hi-C comparisons.

Although developed for Hi-C data analysis, BART3D can also be applied to other 3D-genome data, such as ChIA-PET [38], HiChIP [14], and PLAC-seq [15]. When analyzing HiChIP or PLAC-seq data using BART3D, one may notice that the ChIP factor tends to appear on the top of the inferred TR list. Because HiChIP/PLAC-seq signals are always enriched at genomic binding sites of the ChIP factor regardless of chromatin interaction changes, the inference of a ChIP factor and its known co-factors should be considered false positives and removed for result interpretation. We plan to account for this effect and develop an extended version for

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.19.258095; this version posted August 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

analyzing HiChIP/PLAC-seq data in the future. In addition, TR inference in BART3D is limited to collected ChIP-seq data, which currently include 918 human TRs and 565 mouse TRs but still grow rapidly and require regular updates and maintenance. Nevertheless, BART3D provides a framework for accurate inference of TRs associated with differential chromatin interactions and has broad applications in making biologically meaningful inferences and generating hypotheses from 3D genome data.

CONCLUSIONS

In conclusion, our results demonstrate that BART3D can infer functional TRs associated with dynamic chromatin interactions from differential Hi-C data, despite the resolution of Hi-C data is lower than most TR binding patterns. BART3D is a useful tool for differential analysis of Hi-C data and can help provide biological insights from 3D genome profiles.

METHODS

Data collection

Hi-C, HiChIP, and ChIP-seq data were collected from NCBI GEO [39] in fastq format. Detailed information including accession numbers of all samples used in this work can be found in Supplementary Tables S1-S3.

Data processing

Hi-C (Supplementary Table S1) and HiChIP (Supplementary Table S3) sequence reads were aligned to the human (hg38) or mouse (mm10) reference genomes and processed using HiC-Pro [23]. Contact matrices were generated at a resolution of 5kb and normalized as described in *Normalization of contact matrices*. ChIP-seq (Supplementary Table S2) reads were aligned to the mouse reference genome (mm10) using BWA [40]. Sam files were then converted into bam files using samtools [41]. MACS2 [42] was used to call peaks under the FDR threshold of 0.05.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.19.258095; this version posted August 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Normalization of contact matrices

Given a Hi-C contact matrix $A = \{a_{ij}\}$, the observed read count a_{ij} represents the interaction frequency between a pair of genomic bins i and j. To account for the negative correlation between the intra-chromosomal interaction frequency and the genomic distance between the bin pair [12], we normalized the contact matrix of each chromosome as follows: for any given genomic distance $d_k = k * r$, where r is the bin size (data resolution), we employed a normalization factor $\bar{S}_{d_k}$ as the average read count across all bin pairs with the same genomic distance d_k in this chromosome, i.e., $\bar{S}_{d_k} = (\sum_{j-i=k} a_{ij})/n$, where *n* is the total number of bin pairs with distance d_k. The read count a_{ij} of the bin pair with distance d_k was normalized by $\bar{S}_{d_k}$ as $a'_{ij} = a_{ij}/\bar{S}_{d_k}$. The matrix A was normalized into $A' = \{a'_{ij}\}$ for each chromosome.

Detection of differential chromatin interactions

We denoted the normalized Hi-C contact matrices as $T = \{t_{ij}\}$ for the Treatment condition and $C = \{c_{ij}\}$ for the Control condition, respectively, and $\mathcal{B} = \{1, 2, \ldots, \lfloor l/r \rfloor\}$ representing all equal-sized non-overlapping bins within a chromosome, where l is the length of the chromosome and r is the bin size. For a given genomic region $x \in \mathcal{B}$ and a pre-defined range of genomic distance L, the interaction frequencies between x and its flanking regions with genomic distance up to L were collected, as $IT = \{t_{xk}\}$ from T and $IC = \{c_{xk}\}$ from C, respectively, where $k \in \mathcal{B}, x - L/r \leq k \leq x + L/r$. The differential chromatin interaction (DCI) score at x was calculated as the paired-sample *t*-test statistic between the two arrays of interaction frequencies IT and IC as follows:

$$d_{xk} = t_{xk} - c_{xk},$$

$$DCI = \frac{\bar{d}}{s_d/\sqrt{n}}$$

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.19.258095; this version posted August 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

where d_{xk} is the difference in interactions between each paired element in IT and IC; $\bar{d}$ and s_d are the mean and standard deviation of $\{d_{xk}\}$, respectively; n is the length of each array. Here we have $n = 2L/r + 1$.

Inference of TRs associated with differential chromatin interactions

We used previously curated union DNaseI hypersensitive sites (UDHS), which include 2,723,010 unique non-overlapping DNase-seq peaks for human and 1,529,448 for mouse, to represent all putative cis-regulatory elements (CREs) in the genome [26]. A genome-wide DCI profile was generated by calculating the DCI score of every bin across each chromosome. The DCI profile was mapped to UDHS such that the score for each candidate CRE is set to be equal to the DCI score of the genomic bin where the CRE is located.

We used the BART algorithm to infer TRs associated with differential chromatin interactions [20]. The analysis was done twice, for inferring TRs associated with increased and decreased chromatin interactions separately. For increased chromatin interactions, we ranked all CREs decreasingly by their scores, i.e., CREs with high positive scores would be ranked at the front. We calculated an association score between the CRE profile and each TR binding profile for all ChIP-seq datasets. The association score is defined as the area under the ROC curve (AUC) using the DCI score on CRE as the predictor for TR binding, set as a binary value indicating whether the CRE is overlapped with a peak of that TR from the ChIP-seq dataset. To account for multiple ChIP-seq datasets for the same TR, the Wilcoxon rank-sum test was then applied to assess each TR's significance by comparing the association scores from all ChIP-seq data for this TR with those from all other ChIP-seq datasets, and a background model was used to detect the specificity of each TR. A series of quantification scores with statistical assessments were included for a final ranked list of inferred TRs. For decreased chromatin interactions, the CRE profile was flipped, so that the lowest DCI scores representing CREs the most decreased

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.19.258095; this version posted August 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

chromatin interactions are ranked at the front, and the BART analysis was then performed in the same way.

Determination of default genomic distance parameter

Using different genomic distance range parameters might lead to different TR inference results, because the acting range of different TRs vary a lot. In practice, users may try different distance parameters for exploratory studies. To set an appropriate default value for this parameter, we applied BART3D on a series of Hi-C datasets (Supplementary Table S1) comparing the wide type with perturbation (deletion or activation) of different TRs using different genomic distance ranges, i.e., 50 kb, 100 kb, 200 kb, 500 kb, and 1Mb. We compared the rank of the perturbation target factor in the BART3D results across different genomic distance ranges, and found that 200 kb is where most perturbation factors were ranked on top (Fig. 5). Therefore, we set 200 kb as the default value for the genomic distance range parameter.

Availability of data and materials

Implemented in Python, BART3D package with source code is freely available at https://github.com/zanglab/bart3d

Acknowledgements

The authors would like to thank Dr. Panagiotis Ntziachristos for helpful discussions and members of the Zang Lab for critical reading of the manuscript and testing the software. This work was supported by the US National Institutes of Health [R35GM133712, K22CA204439] and a Phi Beta Psi Sorority Research Grant to C.Z.

Authors' contributions

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.19.258095; this version posted August 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

C.Z. conceived the project and designed the method. Z.W. and Y.Z. implemented the method and performed the analysis. Z.W. and C.Z. wrote the manuscript.

REFERENCES

1. Bonev B, Cavalli G. Organization and function of the 3D genome. Nature Reviews Genetics. Nature Publishing Group; 2016;17:661–78.

2. Yu M, Ren B. The Three-Dimensional Organization of Mammalian Genomes. Annu. Rev. Cell Dev. Biol. Annual Reviews; 2017;33:265–89.

3. Gorkin DU, Leung D, Ren B. The 3D Genome in Transcriptional Regulation and Pluripotency. Cell Stem Cell. Cell Press; 2014;14:762–75.

4. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature. Nature Publishing Group; 2012;485:376–80.

5. Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. Cell. Cell Press; 2014;159:1665–80.

6. Schmitt AD, Hu M, Jung I, Xu Z, Qiu Y, Tan CL, et al. A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. Cell Reports. Cell Press; 2016;17:2042–59.

7. Zheng H, Xie W. The role of 3D genome organization in development and cell differentiation. Nat. Rev. Mol. Cell Biol. Nature Publishing Group; 2019;20:535–50.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.19.258095; this version posted August 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

8. Dixon JR, Jung I, Selvaraj S, Shen Y, Antosiewicz-Bourget JE, Lee AY, et al. Chromatin architecture reorganization during stem cell differentiation. Nature. Nature Publishing Group; 2015;518:331–6.

9. Stadhouders R, Vidal E, Serra F, Di Stefano B, Le Dily F, Quilez J, et al. Transcription factors orchestrate dynamic interplay between genome topology and gene regulation during cell reprogramming. Nat. Genet. Nature Publishing Group; 2018;50:238–49.

10. Kim K-D, Tanizawa H, Iwasaki O, Noma K-I. Transcription factors mediate condensin recruitment and global chromosomal organization in fission yeast. Nat. Genet. Nature Publishing Group; 2016;48:1242–52.

11. van Steensel B, Furlong EEM. The role of transcription in shaping the spatial organization of the genome. Nat. Rev. Mol. Cell Biol. Nature Publishing Group; 2019;20:327–37.

12. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, et al. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. Science. American Association for the Advancement of Science; 2009;326:289–93.

13. Doyle B, Fudenberg G, Imakaev M, Mirny LA. Chromatin Loops as Allosteric Modulators of Enhancer-Promoter Interactions. PLoS Comput Biol. Public Library of Science; 2014;10:e1003867.

14. Mumbach MR, Rubin AJ, Flynn RA, Dai C, Khavari PA, Greenleaf WJ, et al. HiChIP: efficient and sensitive analysis of protein-directed genome architecture. Nature Methods. Nature Publishing Group; 2016;13:919–22.

15. Fang R, Yu M, Li G, Chee S, Liu T, Schmitt AD, et al. Mapping of long-range chromatin interactions by proximity ligation-assisted ChIP-seq. Cell Research. Nature Publishing Group;

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.19.258095; this version posted August 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

2016;26:1345–8.

16. Lun ATL, Smyth GK. diffHic: a Bioconductor package to detect differential genomic interactions in Hi-C data. BMC Bioinformatics. BioMed Central; 2015;16:258.

17. Djekidel MN, Chen Y, Zhang MQ. FIND: difFerential chromatin INteractions Detection using a spatial Poisson process. Genome Res. Cold Spring Harbor Lab; 2018;28:412–22.

18. Stansfield JC, Cresswell KG, Vladimirov VI, Dozmorov MG. HiCcompare: an R-package for joint normalization and comparison of HI-C datasets. BMC Bioinformatics. BioMed Central; 2018;19:279.

19. Ardakany AR, Ay F, Lonardi S. Selfish: discovery of differential chromatin interactions via a self-similarity measure. Bioinformatics. 2019;35:i145–53.

20. Wang Z, Civelek M, Miller CL, Sheffield NC, Guertin MJ, Zang C. BART: a transcription factor prediction tool with query gene sets or epigenomic profiles. Bioinformatics. 2018;34:2867–9.

21. Qin Q, Fan J, Zheng R, Wan C, Mei S, Wu Q, et al. Lisa: inferring transcriptional regulators through integrative modeling of public chromatin accessibility and ChIP-seq data. Genome Biol. BioMed Central; 2020;21:32.

22. Oki S, Ohta T, Shioi G, Hatanaka H, Ogasawara O, Okuda Y, et al. ChIP-Atlas: a data-mining suite powered by full integration of public ChIP-seq data. EMBO Rep. John Wiley & Sons, Ltd; 2018;19:e46255.

23. Servant N, Varoquaux N, Lajoie BR, Viara E, Chen C-J, Vert J-P, et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol. BioMed Central; 2015;16:259.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.19.258095; this version posted August 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

24. Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, et al. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. Cell Systems. Cell Press; 2016;3:95–8.

25. Abdennur N, Mirny LA. Cooler: scalable storage for Hi-C data and other genomically labeled arrays. Bioinformatics. 2019;36:311–6.

26. Wang S, Zang C, Xiao T, Fan J, Mei S, Qin Q, et al. Modeling cis-regulation with a compendium of genome-wide histone H3K27ac profiles. Genome Res. Cold Spring Harbor Lab; 2016;26:1417–29.

27. Rosa-Garrido M, Chapski DJ, Schmitt AD, Kimball TH, Karbassi E, Monte E, et al. High-Resolution Mapping of Chromatin Conformation in Cardiac Myocytes Reveals Structural Remodeling of the Epigenome in Heart Failure. Circulation. 2017;136:1613–25.

28. Canzio D, Nwakeze CL, Horta A, Rajkumar SM, Coffey EL, Duffy EE, et al. Antisense lncRNA Transcription Mediates DNA Demethylation to Drive Stochastic Protocadherin α Promoter Choice. Cell. 2019;177:639–653.e15.

29. Fudenberg G, Imakaev M, Lu C, Goloborodko A, Abdennur N, Mirny LA. Formation of Chromosomal Domains by Loop Extrusion. Cell Reports. 2016;15:2038–49.

30. Kim YH, Marhon SA, Zhang Y, Steger DJ, Won K-J, Lazar MA. Rev-erbα dynamically modulates chromatin looping to control circadian gene transcription. Science. American Association for the Advancement of Science; 2018;359:1274–7.

31. Jansz N, Keniry A, Trussart M, Bildsoe H, Beck T, Tonks ID, et al. Smchd1 regulates long-range chromatin interactions on the inactive X chromosome and at Hox clusters. Nat. Struct. Mol. Biol. Nature Publishing Group; 2018;25:766–77.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.19.258095; this version posted August 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

32. Peters JM, Tedeschi A, Schmitz J. The cohesin complex and its roles in chromosome biology. Genes Dev. Cold Spring Harbor Lab; 2008;22:3089–114.

33. Fontaine C, Dubois G, Duguay Y, Helledie T, Vu-Dac N, Gervois P, et al. The Orphan Nuclear Receptor Rev-Erbα Is a Peroxisome Proliferator-activated Receptor (PPAR) γ Target Gene and Promotes PPARγ-induced Adipocyte Differentiation. J. Biol. Chem. American Society for Biochemistry and Molecular Biology; 2003;278:37672–80.

34. Chamouton J, Latruffe N. PPARα/HNF4α Interplay on Diversified Responsive Elements. Relevance in the Regulation of Liver Peroxisomal Fatty Acid Catabolism. CDM. 2012;13:1436–53.

35. Li X, Xu M, Wang F, Kohan AB, Haas MK, Yang Q, et al. Apolipoprotein A-IV Reduces Hepatic Gluconeogenesis through Nuclear Receptor NR1D1. J. Biol. Chem. American Society for Biochemistry and Molecular Biology; 2014;289:2396–404.

36. Yin L, Wu N, Curtin JC, Qatanani M, Szwergold NR, Reid RA, et al. Rev-erb , a Heme Sensor That Coordinates Metabolic and Circadian Pathways. Science. American Association for the Advancement of Science; 2007;318:1786–9.

37. Jansz N, Nesterova T, Keniry A, Iminitoff M, Hickey PF, Pintacuda G, et al. Smchd1 Targeting to the Inactive X Is Dependent on the Xist-HnrnpK-PRC1 Pathway. Cell Reports. 2018;25:1912–9.

38. Fullwood MJ, Liu MH, Pan YF, Liu J, Xu H, Bin Mohamed Y, et al. An oestrogen-receptor-alpha-bound human chromatin interactome. Nature. 2009;462:58–64.

39. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Res. Narnia; 2013;41:D991–5.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.19.258095; this version posted August 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

40. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. Oxford University Press; 2009;25:1754–60.

41. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078–9.

42. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based Analysis of ChIP-Seq (MACS). Genome Biol. BioMed Central; 2008;9:R137.

Figure 1

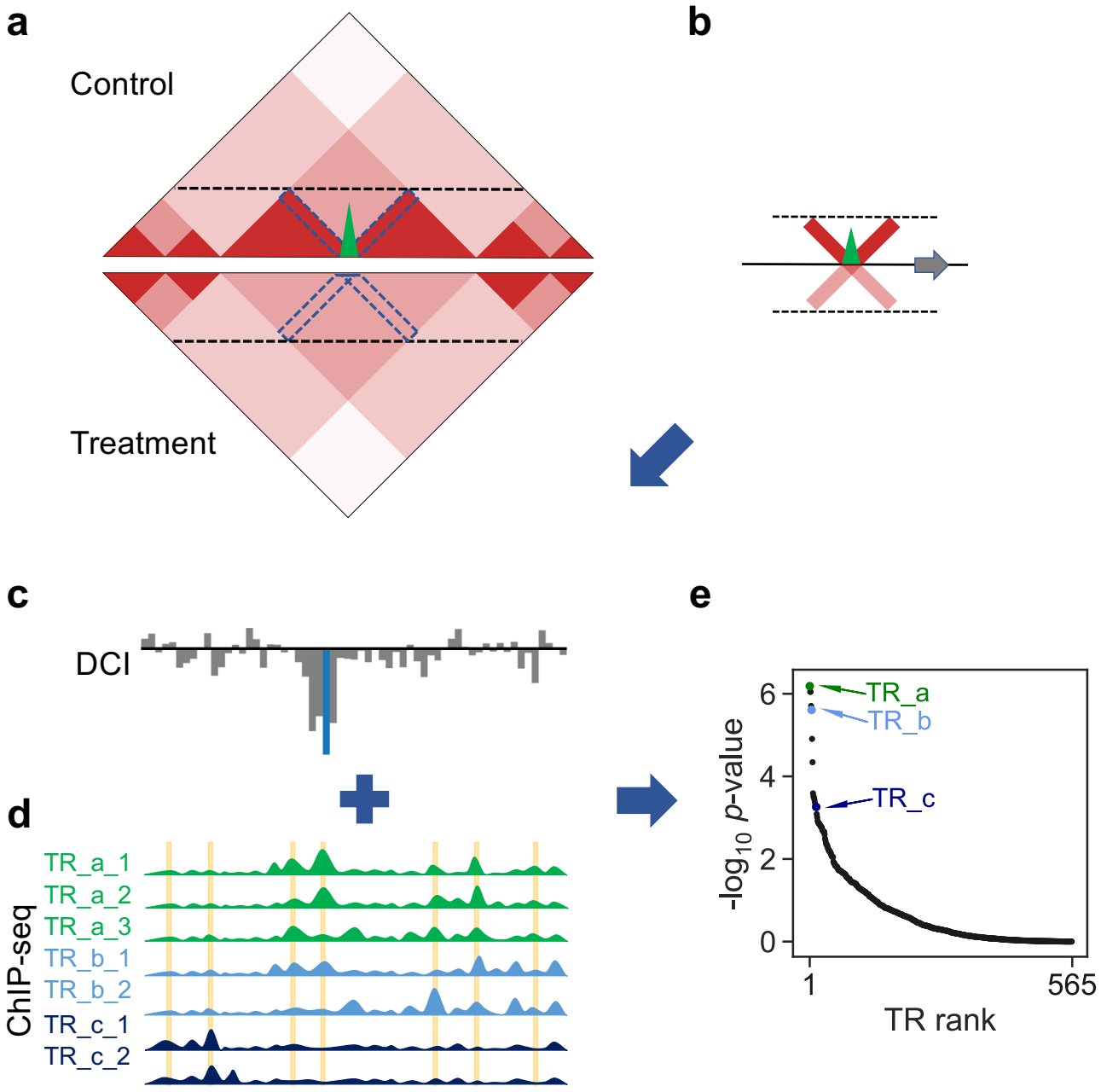


Figure 1. BART3D workflow. BART3D takes as input two Hi-C contact matrices (Control and Treatment) (**a**), scans each chromosome to calculate the DCI score at every bin by comparing the interaction counts (blue dashed boxes at 45-degree angle) within a certain distance boundary (black horizontal dashed lines) between the two conditions (**b**), and derives a DCI profile (**c**). The BART algorithm is then applied to associate the DCI profile with a large collection of public TR ChIP-seq data (**d**) for TR inference analysis. BART3D output is two ranked lists of all TRs associated with increase or decrease of chromatin interactions (**e**).

Figure 2

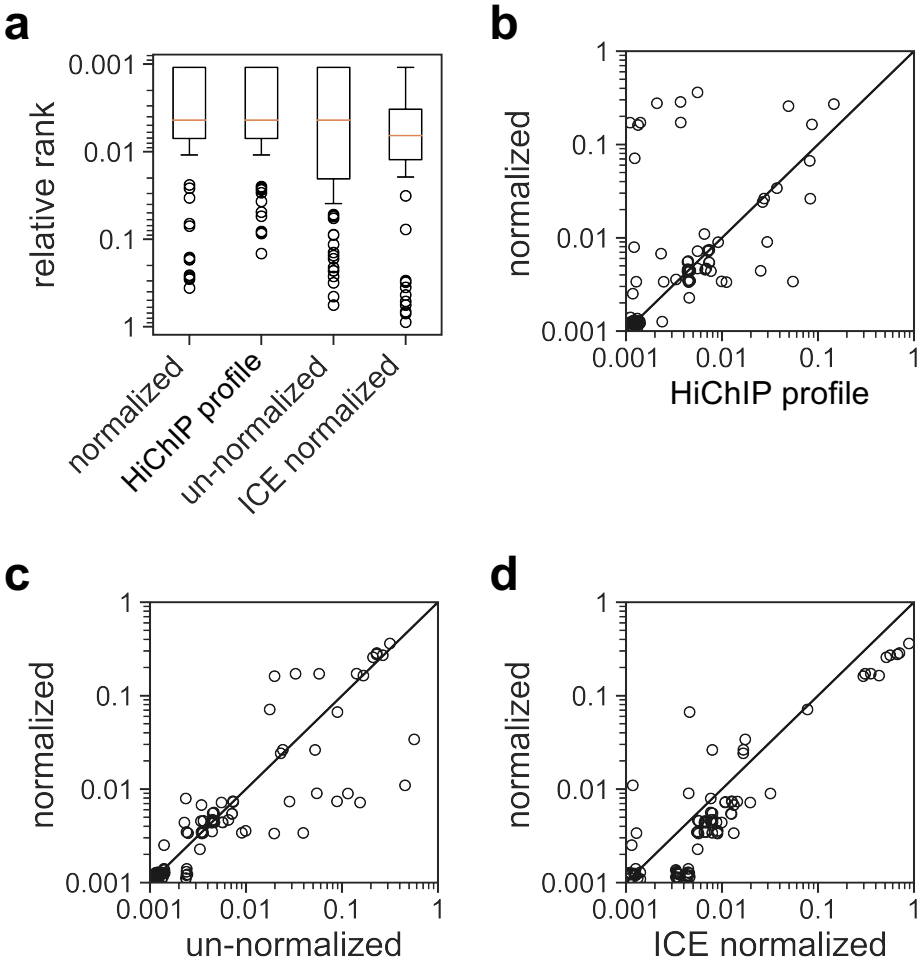


Figure 2. Comparison of different normalization methods.

(**a**) BART results of target TRs from 84 HiChIP datasets using different normalization methods. For distance-based normalized (labeled as “normalized”), unnormalized, and ICE normalized, BART was applied to a genomic region profile scored by summarizing the interaction frequencies of each 5kb bin to its flanking bins within 500kb. For the HiChIP profile, BART was applied to the HiChIP sequence read bam file (as positive control). Relative rank represents the rank of the target TR divided by the total number of TRs in the BART library. Center line in the box represents median.

(**b-d**) Comparison of the relative ranks of target TRs in the BART results generated from distance-based normalization against other methods: (**b**) HiChIP profile, (**c**) unnormalized and (**d**) ICE normalization. Each dot represents a dataset from the 84 HiChIP samples. More dots located below the diagonal line indicates that distance-based normalization (y-axis) yields to higher rank in the BART result.

Figure 3

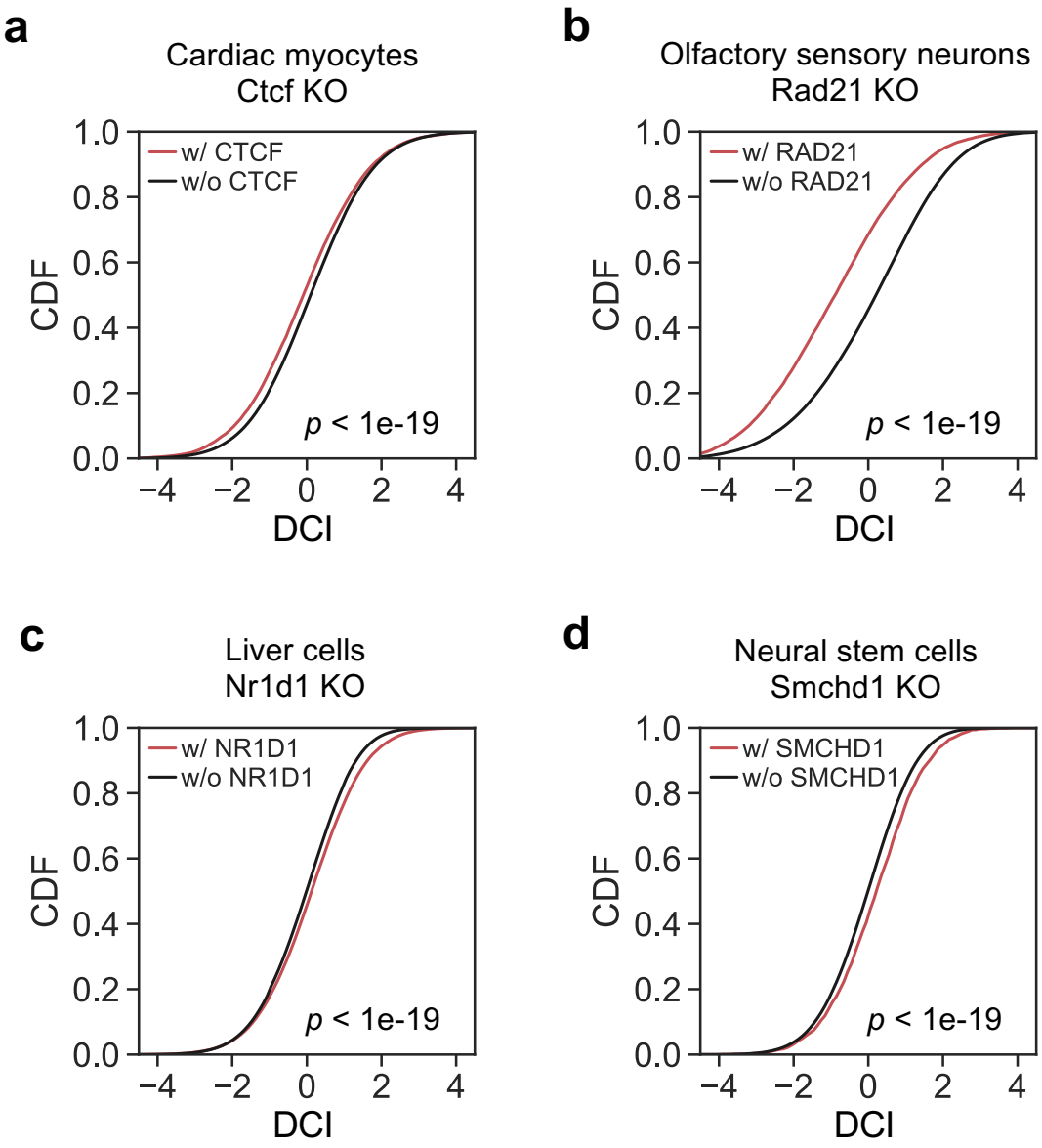


Figure 3. DCI scores at binding sites of KO TRs show significant deviations from the genomic background.

(**a-d**) Cumulative distributions of DCI scores for genomic regions with (red) and without (black) binding sites of KO TRs, including Ctcf KO in cardiac myocytes (**a**), Rad21 KO in olfactory sensory neurons (**b**), Nr1d1 KO in liver cells (**c**), and Smchd1 KO in neural stem cells (**d**), in mouse cell samples. DCI scores were calculated for each 5kb bin by comparing the normalized contact frequencies with its +/-500kb flanking regions. P-values were calculated by Wilcoxon rank-sum test.

Figure 4

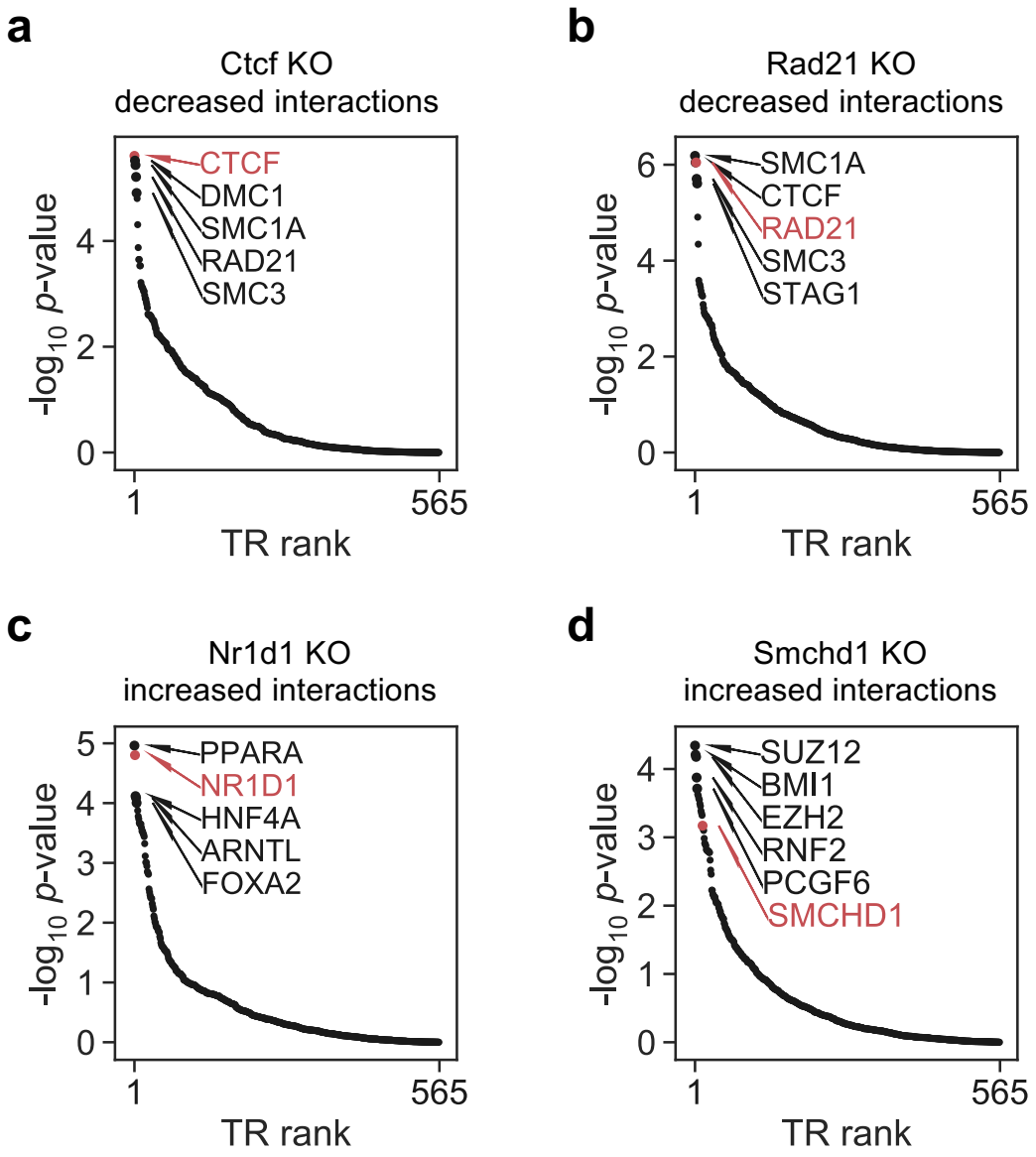


Figure 4. BART3D infers TRs that induce differential chromatin interactions from TR knockout Hi-C data.

(**a-d**) BART3D results focusing on decreased (**a,b**) or increased (**c,d**) chromatin interactions from differential Hi-C datasets in TR KO mouse samples: Ctcf KO in cardiac myocytes (**a**), Rad21 KO in olfactory sensory neurons (**b**), Nr1d1 KO in liver cells (**c**), and Smchd1 KO in neural stem cells (**d**). P-value scores were calculated from rank sum using the null hypothesis under the Irwin-Hall distribution. Top ranked TRs were labeled, and the KO TRs were marked in red.

Figure 5

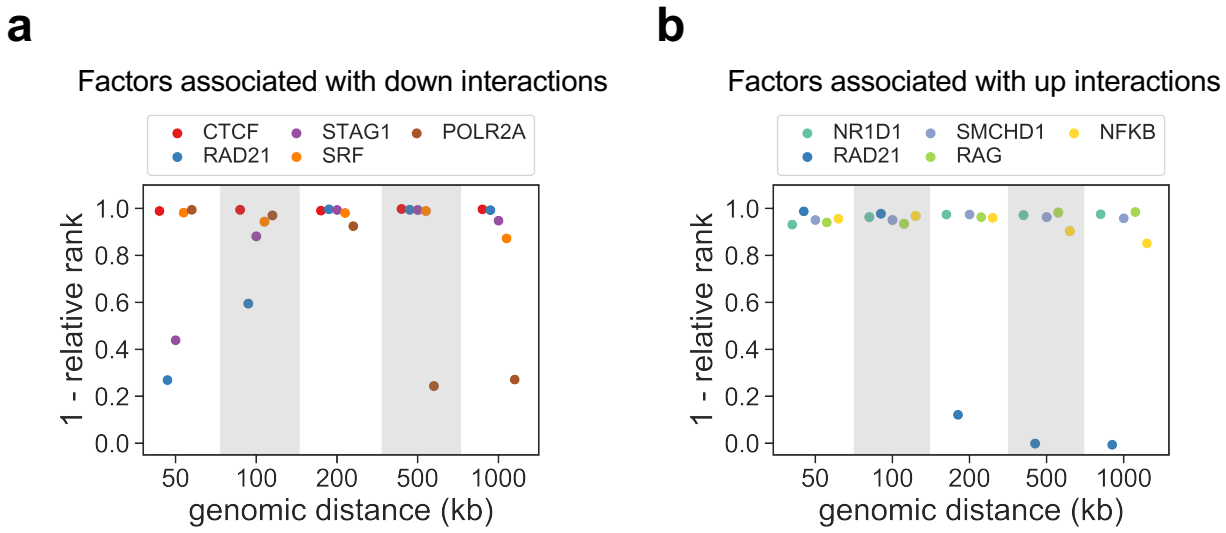
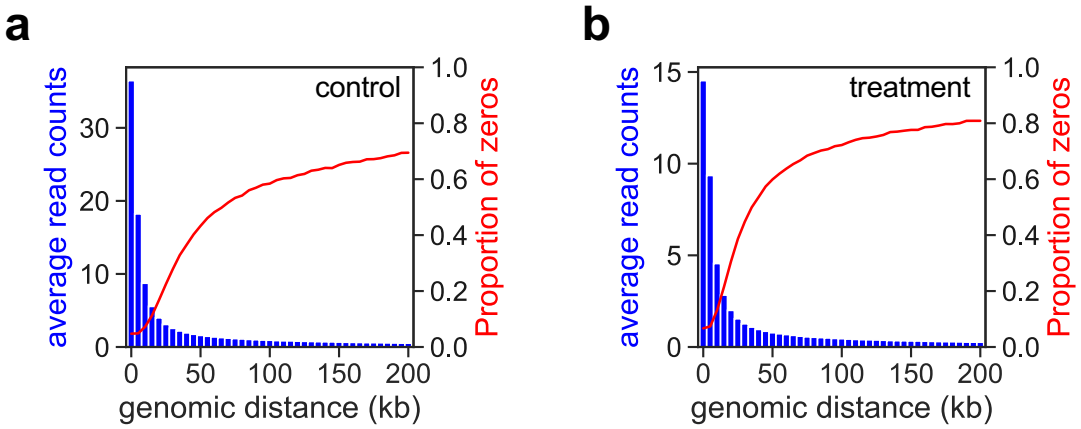


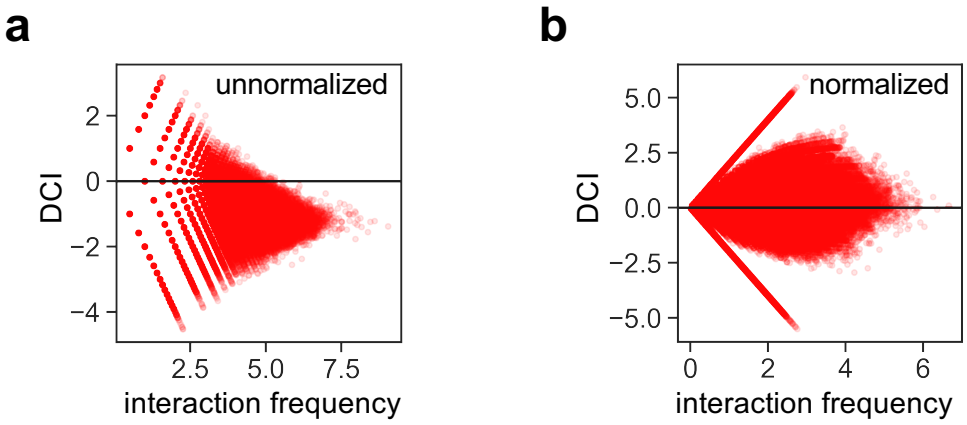
Figure 5. BART3D results on dynamic Hi-C datasets with TR perturbation under different genomic distance settings. The data were separated for TRs associated with decreased (**a**) and increased (**b**) chromatin interactions. The 1 - relative rank of the perturbed TR in BART3D results were shown for each dataset. Higher scores correspond to higher ranked TRs.

Supplementary Figure S1



Supplementary Figure S1. Hi-C read count negatively correlates with genomic distance between bin pairs. Average read counts and percentage of zeros in all bin pairs at the same genomic distance in two Hi-C datasets. (a) control: GSM2790405; (b) treatment: GSM2790406.

Supplementary Figure S2



Supplementary Figure S2. Effect of normalization. MA plots of averaged interaction frequency (x-axis) and differential chromatin interaction (DCI, y-axis) with unnormalized (a) and normalized (b) Hi-C contact matrices between treatment (GSM2790406) and control (GSM2790405).