

A computational framework for identifying the transcription factors involved in enhancer-promoter loop formation

Li Liu,¹ Li-Rong Zhang,¹ Fu-Ying Dao,² Yan-Chao Yang,¹ and Hao Lin²

¹Laboratory of Theoretical Biophysics, School of Physical Science and Technology, Inner Mongolia University, Hohhot 010021, China; ²School of Life Science and Technology, Center for Informational Biology, University of Electronic Science and Technology of China, Chengdu 610054, China

The pairwise interaction between transcription factors (TFs) plays an important role in enhancer-promoter loop formation. Although thousands of TFs in the human genome have been found, only a few TF pairs have been demonstrated to be related to loop formation. It is still a challenge to determine which TF pairs could be involved in the enhancer-promoter regulation network. This work describes a computational framework to identify TF pairs in enhancer-promoter regulation. By integrating different levels of data derived from Promoter Capture Hi-C, chromatin immunoprecipitation sequencing (ChIP-seq) of histone marks, RNA-seq, protein-protein interaction (PPI), and TF motif, we identified 361 significant TF pairs and constructed a TF interaction network. From the network, we found several hub-TFs, which may have important roles in the regulation of long-range interactions. Our studies extended TF pairs identified in other experimental and computational approaches. These findings will help the further study of long-range interactions between enhancers and promoters.

INTRODUCTION

The study of gene regulation has shifted the focus from linear genome to 3D genome in recent years. Experimental methods such as Hi-C,¹ ChIA-PET,² HiChIP,³ and Capture Hi-C⁴ have revealed that 3D chromatin architecture plays an important role in gene transcriptional regulation. The remote enhancer may activate the target promoter through physical contacts, whereby the target gene can be expressed in different cell types or tissues and in stress response to environmental changes. There is a general consensus that sequence-specific transcription factors bind at enhancers and then recruit more cofactor complexes to mediate communication with the target promoters.⁵ Considerable evidence demonstrates that several transcription factors such as CTCF, YY1, ER α , ZNF143, EKLF, and GATA1 can facilitate enhancer-promoter interaction. The zinc-coordinating proteins CTCF or YY1 binding at enhancers and promoters can form homodimers and thus facilitate loop formation.⁶ The ER α protein often binds to regulatory DNA elements distant from gene promoters and interacts with other factors that bind to promoters, such as FoxA1 and RNAPII, to form chromatin looping structures.⁷ The zinc-finger protein ZNF143 is directly recruited to the promoter

of gene to physically bridge the promoter and CTCF-cohesin cluster.⁸ The transcription factors EKLF,⁹ GATA1,¹⁰ and CTCF¹¹ are required for the physical interaction between the β -globin locus control region (LCR) and the β -globin promoter. Besides these transcription factors that directly bind to the DNA molecular, some cofactors, which indirectly bind to DNA, also play important roles in loop formation. Cohesin, the structural maintenance of chromosomes protein complex, can form a ring-shaped structure that connects two DNA segments and contributes to stabilizing enhancer-promoter interactions.¹² Mediator, a highly conserved transcriptional coactivator, can physically bridge enhancer-bound transcription factors and the promoter-bound proteins. The flexible conformation, which is comprised of variable subunits, is important for Mediator's ability to bind various proteins.¹³

Collectively, the key players in 3D genome architecture, such as CTCF and cohesin complex, have been widely and deeply studied. These studies have demonstrated that the cooperation of multiple protein factors is critical to orchestrate loop formation. However, only a few new protein factors, such as YY1, have been identified over the past couple of years due to technical limitations. Chromatin immunoprecipitation with mass spectrometry (ChIP-MS) provides a way for *de novo*-seeking candidates. The ChIP-MS-based "Chromatin Proteomic Profiling" method has been used to identify proteins associated with genomic regions marked by histones modified at specific lysine residues.¹⁴ To identify proteins that bind to active enhancers and promoters simultaneously, Weintraub et al.⁶ implemented a modified histone ChIP-MS method, which used antibodies directed toward histone modifications H3K27ac and H3K4me3. They provided 26 candidate transcription factors and concluded that YY1 is essential for

Received 13 July 2020; accepted 11 November 2020;
<https://doi.org/10.1016/j.omtn.2020.11.011>

Correspondence: Li-Rong Zhang, Laboratory of Theoretical Biophysics, School of Physical Science and Technology, Inner Mongolia University, Hohhot 010021, China.

E-mail: pyzlr@imu.edu.cn

Correspondence: Hao Lin, School of Life Science and Technology, Center for Informational Biology, University of Electronic Science and Technology of China, Chengdu 610054, China.

E-mail: hlin@uestc.edu.cn



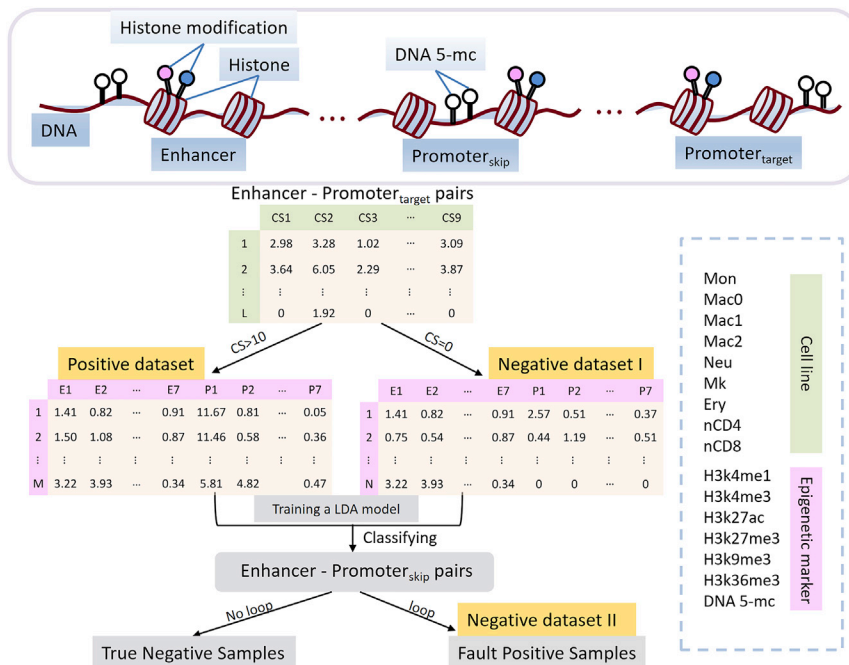


Figure 1. The pipeline for constructing datasets

The pipeline contains three main stages: (1) mapping enhancers and promoters to loop anchors, (2) dividing enhancer-promoter pairs into positive dataset and negative dataset I based on CHI-CAGO scores (CS1-CS9), and (3) training LDA model to select samples for negative dataset II. The promoters that overlap the anchors are defined as promoter_{target}. The promoter_{skip} represents the one that is skipped by loop. The epigenetic markers of enhancer (E1-E7) and promoter (P1-P7) are used to encode an enhancer-promoter pair for the classifier.

Furthermore, several hub-TFs were identified as the important candidate protein factors through the interaction network.

RESULTS

Epigenetic modifications in enhancer-promoter loop

By using our pipeline shown in Figure 1, we obtained 47,987 high-confidence enhancer-promoter loops (positive dataset) and 80,844 enhancer-promoter pairs (negative dataset I)

that didn't form loops in corresponding cell types. The different distributions of epigenetic marks were compared and shown in Figure 2. We found that some active histone modifications, such as H3K27ac, H3K4me1, and H3K36me3, are enriched in positive dataset at both enhancer and promoter regions. Conversely, inhibitive histone modification signals, such as H3K27me3 and H3K9me3, are weak at these regions. These results are consistent with previous findings.^{6,18} Thus, we may conclude that the epigenetic modifications can affect the formation of chromatin loops. It implies that the epigenetic features can be used for the prediction of enhancer-promoter loops.

Based on above analysis, we trained a linear discriminant analysis (LDA) model on positive dataset and negative dataset I using 14 epigenetic modification features and made prediction on 3,964,527 pairs of enhancers and skip genes. As a result, 260,996 false-positive enhancer-promoter pairs were extracted to construct the negative dataset II, which was approximately 5 times the size of the positive dataset. The sampling method used here can help to find the factors other than the distance and epigenetic modifications. First, the distance has been used as the important feature for distinguishing true enhancer-promoter pairs from non-interacting pairs,¹⁸⁻²⁰ because majority enhancer-promoter interactions occurred within the topologically associating domain (TAD).^{6,21-24} We focused on skip genes, because the linear distance between enhancer and skip gene is always less than the one in the positive dataset. Moreover, according to the process of loop extrusion, cohesin translocate along DNA sequences until it encounters loop anchors.²⁵⁻²⁸ However, how did it skip the genes between two anchors? There is still no definitive answer. Using the skip genes as the negative samples will contribute

enhancer-promoter structural interactions. Accumulating evidence suggests that more than 1,000 proteins are thought to contribute to mammalian chromatin structure and its regulation. In the present period, it is difficult to produce a complete catalog through a single assay. We have a limited understanding of the transcription factors involved in long-range interactions. Another strategy is to construct the cooperation network based on available ChIP-seq data or motif of transcription factors. Zhang et al.¹⁵ have collected 84 DNA-binding proteins (DBPs) ChIP-seq and kilobase-resolution Hi-C data in GM12878 and K562 cell lines and employed a Gaussian graphical model (GGM) to identify protein combinations mediating chromatin looping. Duren et al.¹⁶ proposed a model based on motif data from 557 TFs to infer candidate TF-TF pairs that facilitated enhancer-promoter cooperation. Diekidel et al.¹⁷ presented a tool (3CPET) based on ChIA-PET and ChIP-seq data from MCF7 and K562 cell lines to find co-factor complexes involved in maintaining chromatin interactions.

The accumulation of omics data provides an opportunity to systematically investigate the regulation network using a computational strategy. In this work, we developed a new model to predict the TF pairs that are possibly involved in enhancer-promoter loop formation. In this model, the TF motif, TF expression, protein-protein interaction (PPI), and activity status of enhancer and promoter were integrated to construct the interaction network. Differing from previous studies, we used a real non-interactive enhancer-promoter network instead of the random network as background distribution. This approach could produce more reliable results. As a result, we got hundreds of significant TF-TF pairs that could physically interact with each other from nine human primary blood cell types.

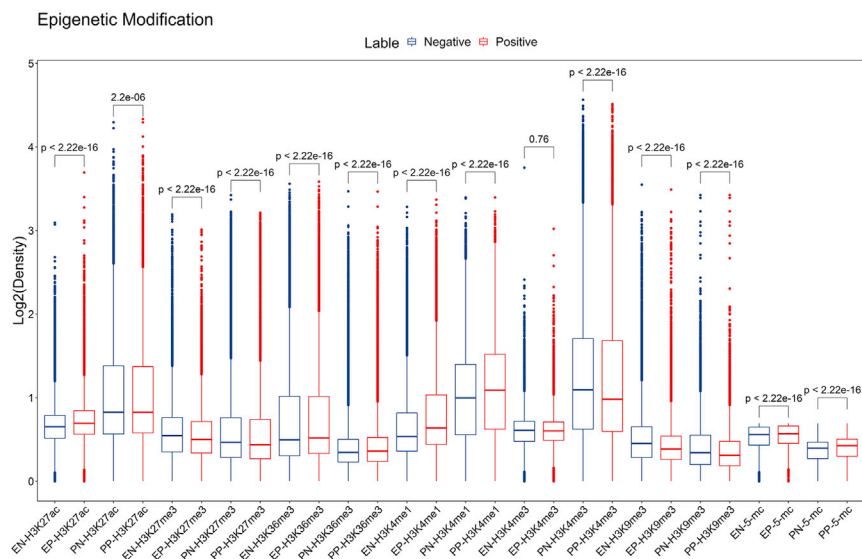


Figure 2. Distribution of epigenetic marks

Each box shows the average density distribution in positive dataset (red) or negative dataset (blue). Each epigenetic mark is calculated individually for enhancer and promoter regions. The prefixes “EN-,” “EP-,” “PN-,” and “PP-” represent “enhancer of negative samples,” “enhancer of positive samples,” “promoter of negative samples,” and “promoter of positive samples,” respectively. p values of Wilcoxon test are displayed over the boxes.

to this question. Second, the false-positive enhancer-promoter pairs predicted by LDA must have the epigenetic marks similar to the true pairs, which is important to exclude the influence of epigenetic modifications.

The discovery of significant TF pairs

On the basis of positive dataset and negative dataset II, the function TF_{ij} defined in Equation 1 was used to select the TF pairs that have the potential to bridge enhancer and promoter. To reduce random background noise, we only considered the motifs that occurred at least 3 times within the 2.5 kb region of enhancer or promoter. In addition, we restricted $RPKM \geq 1$ to ensure that the proteins do expressed in the cell. Using this approach, we got 1,158 TF pairs involved 219 TFs. These TF pairs can be separated into three groups (positive preference, no preference, negative preference) based on the Uscore (Figure 3A). The positive preference group includes 361 TF pairs whose occurrence frequency in positive dataset is higher than that in negative dataset II ($p < 0.01$, Table S1). It is noteworthy that the three most significant TF pairs are MYC-BCL2, BCL2-MYC, and BCL2-BCL2. MYC and BCL2 (B cell lymphoma/leukemia gene 2) are both proto-oncogene that involved in apoptosis.^{29,30} We found that both of them are cell-type-specific expression. Intriguingly, they are almost co-expression across the nine blood cells ($r = 0.58$, Figure 3C). Furthermore, the coverage ratio of them in all samples are less than 50% (Figure 3B). These observations suggest that MYC and BCL2 have the potential to participate in loop formation in cell-type-specific manner.

We next ranked the 1,158 TF pairs according to the coverage ratio P_c . The TF pairs whose P_c values are more than the cutoff (0.9) were defined as ubiquitous TF pairs. These ubiquitous TF pairs, such as YY1-YY1 and SMC3-RAD21, could be necessary condition but not sufficient for loop formation (Figure 3B). And, as expected, YY1,

involved in enhancer-promoter contacts.^{6,31} Thus, our observation is consistent with previous conclusion.

Two computational works have focused on TF pairs identification.^{15,16} Thus, we also compared our findings with their results. Zhang et al.¹⁵ identified hundreds of direct and indirect 3D DBP (DNA-binding proteins) interactions. However, only 45 direct interactions were recognized, which is less than our findings. Our model could find 361 direct interactions. Duren et al.¹⁶ identified 53 candidate cooperating TF-TF pairs that one TF binds to the promoter and the other to an enhancer. In these studies, they did not distinguish the binding sites of promoter region from those of enhancer region. Thus, our results are more reliable. The comparison in Figure 3D and Table S1 showed that 9 TF pairs obtained by us appeared in Duren’s findings. There are also 9 TF pairs that are shared between Zhang’s dataset and our dataset. However, there is no overlap between Duren’s dataset and Zhang’s dataset.

Hub-TFs are found by constructing an interaction network

To systematically elucidate the relationships among these interacting TFs, we employed Cytoscape³² to create a TF interaction network (Figure 4). The network contains 361 physical interactions of 141 TFs. Here, we considered the directionality of interaction. The direction was marked by an arrow from enhancer to promoter. Interestingly, the TF interaction network shows asymmetric degree distribution, which is the hallmark feature of scale-free networks. In previous study, Wang et al.³³ generated genome-wide transcription factor binding site (TFBS)-TFBS networks from human Hi-C data. They first observed that the TFBS-TFBS networks followed a scale-free degree distribution. We therefore highlighted the TFs that had been highly connected. By calculating the degree centrality, we identified several hub-TFs (Table 1). We noticed that many hub-TFs involved in the processes of cell proliferation and differentiation such as EP300, MYC, and RELA. Among these hub-TFs, EP300

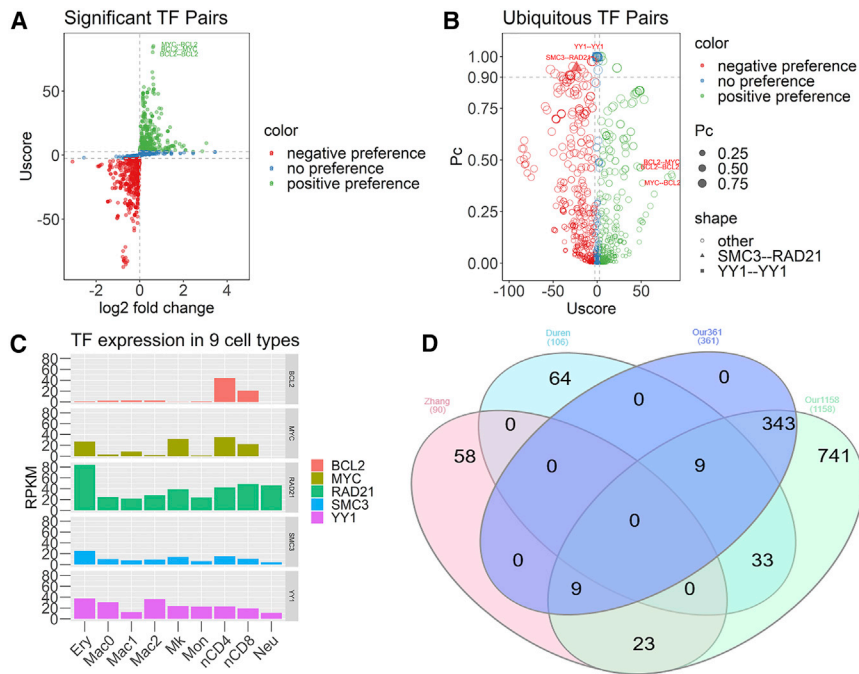


Figure 3. The candidate TF-TF pairs for loop formation

(A) The significant TF pairs identified by UScore. (B) The ubiquitous TF pairs identified by the coverage ratio Pc. (C) Expression of five TFs in nine cell lines. (D) The Venn plot showing the intersection of four TF pair datasets.

DISCUSSION

In this work, we introduced a two-step framework to identify TF pairs associated with enhancer-promoter loops. It started with a large collection of PCHiC datasets and ended at the construction of a TF interaction network. Compared with Hi-C, PCHiC assay offers more high-resolution promoter-centered loop. A large number of cell-type-specific chromatin loops make it more accurate to analyze the effect of epigenetic modifications on loop formation. The key step of the framework is to select a reasonable negative samples dataset. We focused on the skip genes, which were closer to enhancers in linear distance but didn't form loops. Our analyses revealed that the TF pairs

that bound on regulatory regions indeed influenced the loop formation. Our method provided candidate TF pairs by comparing motif distribution between positive samples and negative samples. Using motif as a TF binding site, we could overcome the limitation of unavailability of ChIP-seq data. We also concluded that the well-studied TF pairs, YY1-YY1 and SMC3-RAD21, are necessary but not sufficient for loop formation. One major limitation of this work is that it didn't consider indirect protein interaction. This is an issue for future research to explore. Moreover, gene-gene spatial interaction has been studied based on Hi-C data.^{44,45} Future studies could investigate which TFs drive the clustering of distal genes to be co-regulated. In summary, our computational approach which integrate multi-omics data can help to identify the TFs that may mediate long-range chromatin interactions.

MATERIALS AND METHODS

Data collection

The Promoter Capture Hi-C (PCHiC) data in 17 human primary blood cell types were downloaded from the Open Science Framework (<https://osf.io/u8tzc>).⁴⁶ The processed datasets for 9 of 17 cell types, including histone modification ChIP-seq data, DNA methylation data, and gene quantification RNA-seq data, were obtained from the BLUEPRINT project (ftp://ftp.ebi.ac.uk/pub/databases/blueprint/data/homo_sapiens/GRCh37/).

Generating positive dataset and negative datasets

In terms of the current understanding, the occupancies of transcription factors on *cis*-regulatory elements are restricted by chromatin epigenetic modifications and TF binding sites. In this study, we aim to find the candidate TF pairs based on TF binding information.

(E1A binding protein p300) is the most important one, which is consistent with previous report.¹⁵ EP300 is a histone acetyltransferase.³⁴ The binding sites of EP300 were often used to predict activated enhancers because of the ability to mediate acetylation of histone H3 at "lys-27" (H3K27ac).^{35,36} It can remodel chromatin and interact with many proteins to activate transcription.^{37,38} MYC, a BHLH (basic helix-loop-helix) transcription factor, is a major driver of most human cancers.³⁹ Overexpressed MYC binds to virtually all active promoters within a cell and modulates the expression of distinct subsets of genes.⁴⁰ RELA belongs to a family of transcription factors nuclear factor κ B (NF- κ B) complex (NF- κ from B cells). Beside its activity as a direct transcriptional activator, it is also able to modulate promoter accessibility to transcription factors and thereby indirectly regulate gene expression.⁴¹ Collectively, the hub-TFs in the network play a key role in transcriptional regulation and the abnormal expression of them could induce tumorigenesis.^{29,42,43} Our study suggested that they may also contribute to loop formation.

To inspect whether the hub-TFs have the asymmetric distribution between enhancers and promoters, we calculated the in-degree centrality and out-degree centrality for each node (Table 1). We found no distinct preferences in different regions. Furthermore, we compared enhancer-TFs with promoter-TFs. We found that majority of significant TFs occupy either enhancer region or promoter region (Figure 5; Table S2). Evidence has demonstrated that homodimers, such as CTCF-CTCF and YY1-YY1, play important roles in chromatin loop structure.⁶ Similarly, the network also provided 21 significant TFs whose binding sites coincide with enhancers and target promoters (Table S2). These TFs are capable of forming homodimers to participate in enhancer-promoter loop formation.

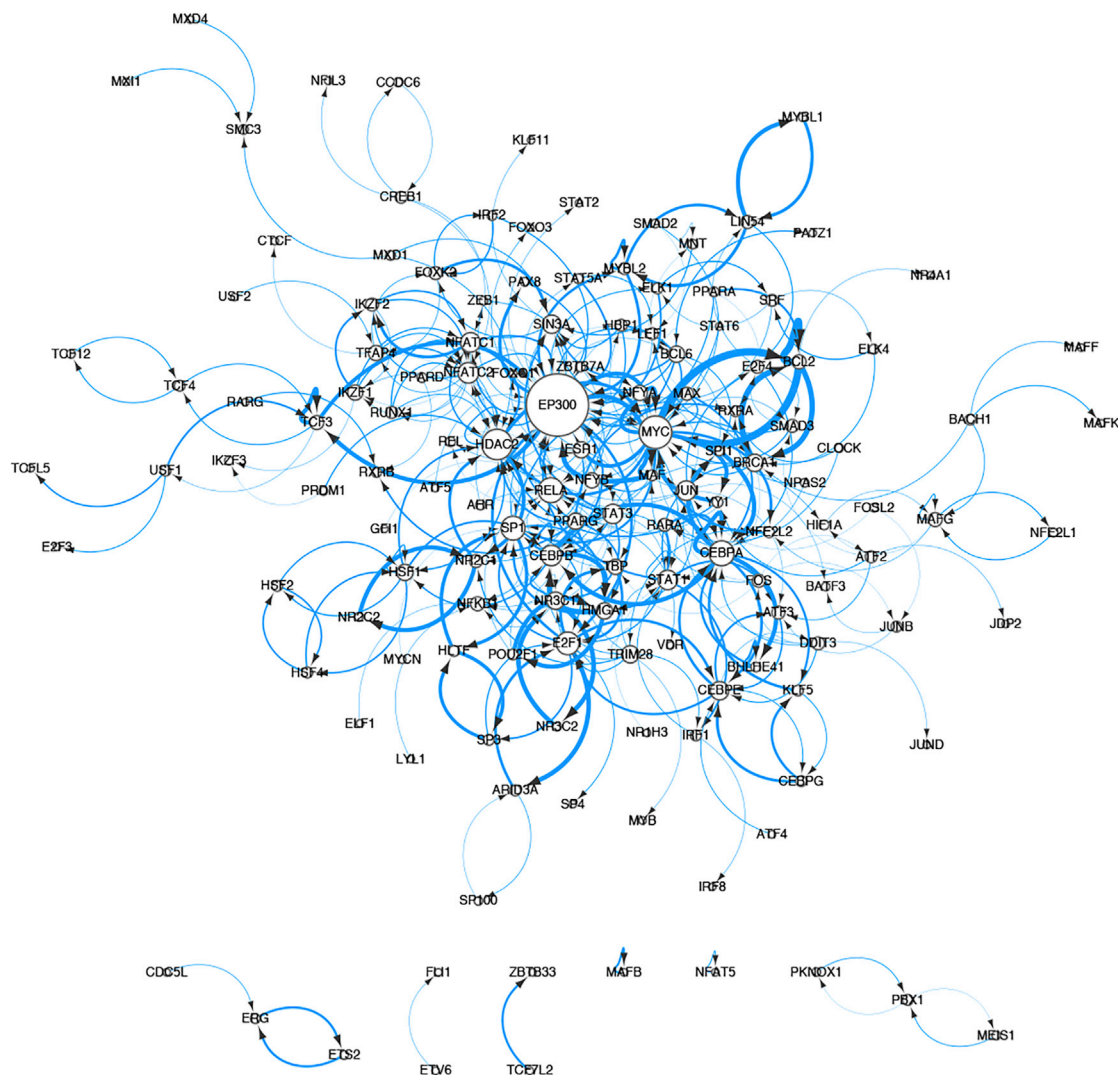


Figure 4. TF interaction network

The directed graph contains 141 nodes and 361 edges. Each node represents a TF, which bound on either promoter region or enhancer region. Each edge indicates a physical interaction between two TFs and the direction of arrow is from enhancer to promoter. The size of circle is matched to Uscore.

Thus, a key step is to construct a negative samples dataset that has been in the same epigenetic modifications state with the positive samples.

The PCHiC dataset gives a list of chromatin interactions with CHiCAGO scores ≥ 5 in at least one cell type.⁴⁷ At first, we mapped enhancers and target gene promoters to the two interaction fragments, respectively. When filtering out the interactions whose either side overlap with more than one enhancer or promoter, we got the one-to-one enhancer-promoter pairs for the following analyses. There are different CHiCAGO scores for the same enhancer-promoter pair in different cell types. We merged the data from nine cell types (Neu, nCD8, nCD4, Mon, Mk, Mac0, Mac1, Mac2, Ery)

and selected the interactions with CHiCAGO scores ≥ 10 as positive dataset, the interactions with CHiCAGO scores = 0 as negative dataset I. Then each enhancer-promoter pair was represented as a 14-dimension feature vector. Each dimension corresponded to one of the 14 epigenetic modifications data. For the positive dataset and negative dataset I, we hypothesized that the different epigenetic modifications directly affect loop formation. Based on this assumption, we trained a linear discriminant analysis (LDA) model on the positive dataset and negative dataset I. We also compared the LDA model with another two algorithms (logistic regression and Naive Bayes). Results were listed in Table S3. Our aim is to select the false positive samples to construct negative dataset II. Due to high specificity produced by LDA mode, we selected LDA as prediction algorithm.

Table 1. The top six hub-TFs

Hub-TF	Degree centrality	Out-degree centrality	In-degree centrality
EP300	51	24	27
MYC	25	12	13
HDAC2	23	6	17
CEBPA	18	8	10
RELA	18	3	15
SP1	16	7	9

To minimize the effect of enhancer-promoter distance and epigenetic modifications, we need another negative dataset. At first, we defined the genes between enhancers and target genes as “skip” genes. Based on the loop extrusion model,^{25,48} these skip genes are good background for our study. Subsequently, we grouped the skip genes into two categories using above LDA model. Those skip genes which were predicted as positive samples were regarded as the negative dataset II. The process for datasets construction were shown in Figure 1. The three datasets are provided in the Data S1.

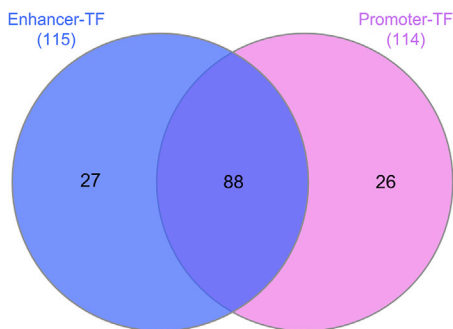
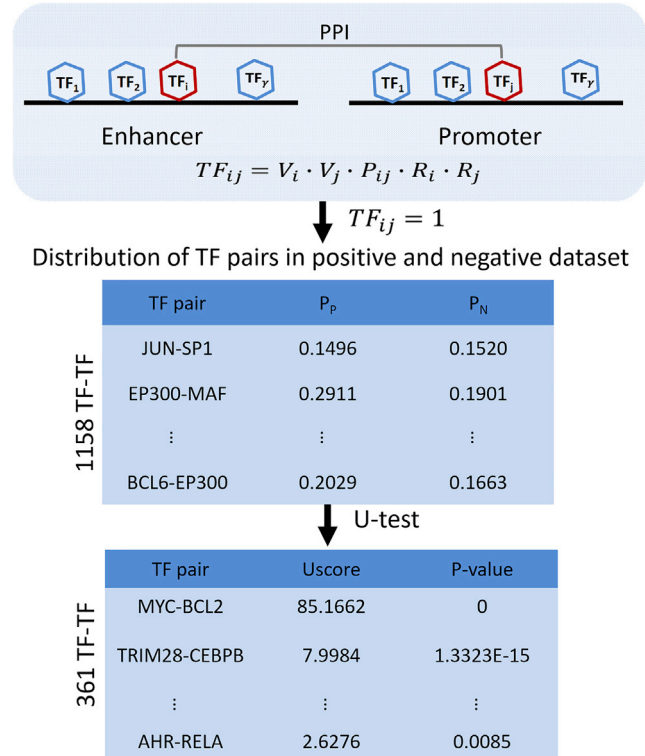
Identifying significant TF pairs

To identify TF-TF pairs that facilitate enhancer-promoter communication, we set up a framework based on the positive dataset and negative dataset II (Figure 6). At first, we collected 705 TFs (including 8,785 motifs) from JASPAR,⁴⁹ TRANSFAC,⁵⁰ UniPROBE,⁵¹ Taipale,⁵² and HOMER⁵³ database. Then we got 327 TFs after overlapping with protein-protein interaction (PPI) database. We defined a function to select TF pairs that have the potential to bridge enhancer and promoter:

$$TF_{ij} = V_i \times V_j \times P_{ij} \times R_i \times R_j \quad (\text{Equation 1})$$

where

$$V_i = \begin{cases} 1 & \text{if the motifs of } TF_i \text{ occurring at least 3 times} \\ 0 & \text{if the motifs of } TF_i \text{ occurring less than 3 times} \end{cases} \quad (\text{Equation 2})$$

**Figure 5. Venn diagram of enhancer-TFs and promoter-TFs****Figure 6. Workflow for identifying significant TF pairs**

In the beginning, all of the TF pairs acted as candidates. After hierarchical screening (PPI, TF_{ij} , and U-test), we identified 361 significant TF pairs. P_p and P_N represent the frequencies of TF pairs occurring in positive dataset and negative dataset II, respectively. U-score and p values represent the significance level of difference.

$$V_j = \begin{cases} 1 & \text{if the motifs of } TF_j \text{ occurring at least 3 times} \\ 0 & \text{if the motifs of } TF_j \text{ occurring less than 3 times} \end{cases} \quad (\text{Equation 3})$$

$$P_{ij} = \begin{cases} 1 & \text{if there is PPI between } TF_i \text{ and } TF_j \\ 0 & \text{if there is no PPI between } TF_i \text{ and } TF_j \end{cases} \quad (\text{Equation 4})$$

$$R_i = \begin{cases} 1 & \text{if RPKM of } TF_i \text{ is greater than or equal to 1} \\ 0 & \text{if RPKM of } TF_i \text{ is smaller than 1} \end{cases} \quad (\text{Equation 5})$$

$$R_j = \begin{cases} 1 & \text{if RPKM of } TF_j \text{ is greater than or equal to 1} \\ 0 & \text{if RPKM of } TF_j \text{ is smaller than 1} \end{cases} \quad (\text{Equation 6})$$

where the TF_{ij} represents a pair of TF, where $1 \leq i \leq 327$ and $1 \leq j \leq 327$. The value of variable TF_{ij} is also binary. When $TF_{ij} = 1$, the TF pair is counted. We calculated the occurrence frequencies of these TF pairs in positive dataset and negative dataset II, respectively.

Then Mann-Whitney U-test was subsequently used to identify the significant TF pairs that prefer to occur in positive dataset:

$$U_{score} = \frac{p_1 - p_2}{\sqrt{p_c \times (1 - p_c) \times \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad (\text{Equation 7})$$

where

$$p_1 = \frac{x_1}{n_1} \quad (\text{Equation 8})$$

$$p_2 = \frac{x_2}{n_2} \quad (\text{Equation 9})$$

$$p_c = \frac{x_1 + x_2}{n_1 + n_2}. \quad (\text{Equation 10})$$

The norms n_1 and n_2 are the numbers of enhancer-promoter pairs in positive dataset and negative dataset II, respectively. The variables x_1 and x_2 are the numbers of enhancer-promoter pairs in which the current TF pair is included. For each TF pair, the statistics U_{score} is calculated to test the significance level.

Data processing

We defined promoters as the regions of upstream 0.5 kb and downstream 2 kb of the transcription start sites (TSS) according to Ensembl v.75 (<http://grch37.ensembl.org>). We defined enhancers on the basis of Regulatory Build⁵⁴ and trimmed or extended them to 2.5 kb while maintaining their original center. To generate enhancer-promoter loops, we mapped promoters and enhancers to PCHi-C loop anchors, respectively. The loops were removed if more than one promoter or enhancer overlapped with the loop anchors. Likewise, the skip gene promoters and corresponding enhancers were sampled as negative samples. The processed bigWig format data of histone modifications (H3K4me1, H3K4me3, H3K9me3, H3K27me3, H3K27ac, H3K36me3) and DNA methylation were used here to quantify the epigenetic modifications. We calculated average epigenetic signals for promoter and enhancer regions. Protein-protein interaction data was downloaded from the BioGrid database.⁵⁵ We extracted a subset that only contained physical interactions of human proteins by searching keywords “taxid:9606” and “physical.” When any TF pair we collected was supported by PPI, it was retained. Next, for each TF, we looked for all the binding sites in the whole genome by using the HOMER sub-program “scanMotifGenomeWide.pl.” Then we counted how many binding sites fell into each promoter region and enhancer region through “intersect” utility of BEDTools.⁵⁶

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.omtn.2020.11.011>.

ACKNOWLEDGMENTS

We acknowledge Zhana Duren for his valuable input during the method development process. This work was supported by grants

from the National Natural Science Foundation of China (61961031, 61962041, 61772119, and 62062053) and the Sichuan Provincial Science Fund for Distinguished Young Scholars (2020DJQ0012).

AUTHOR CONTRIBUTIONS

L.L., H.L., and L.-R.Z. conceived and designed the study. L.L. and H.L. performed computational analyses and wrote the manuscript. F.-Y.D. and Y.-C.Y. contribute to the manuscript revision. All authors read and approved the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

- Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragozy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 289–293.
- Li, G., Ruan, X., Auerbach, R.K., Sandhu, K.S., Zheng, M., Wang, P., Poh, H.M., Goh, Y., Lim, J., Zhang, J., et al. (2012). Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* 148, 84–98.
- Mumbach, M.R., Satpathy, A.T., Boyle, E.A., Dai, C., Gowen, B.G., Cho, S.W., Nguyen, M.L., Rubin, A.J., Granja, J.M., Kazane, K.R., et al. (2017). Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. *Nat. Genet.* 49, 1602–1612.
- Mifsud, B., Tavares-Cadete, F., Young, A.N., Sugar, R., Schoenfelder, S., Ferreira, L., Wingett, S.W., Andrews, S., Grey, W., Ewels, P.A., et al. (2015). Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat. Genet.* 47, 598–606.
- Schoenfelder, S., and Fraser, P. (2019). Long-range enhancer-promoter contacts in gene expression control. *Nat. Rev. Genet.* 20, 437–455.
- Weintraub, A.S., Li, C.H., Zamudio, A.V., Sigova, A.A., Hannett, N.M., Day, D.S., Abraham, B.J., Cohen, M.A., Nabet, B., Buckley, D.L., et al. (2017). YY1 Is a Structural Regulator of Enhancer-Promoter Loops. *Cell* 171, 1573–1588.
- Fullwood, M.J., Liu, M.H., Pan, Y.F., Liu, J., Xu, H., Mohamed, Y.B., Orlov, Y.L., Velkov, S., Ho, A., Mei, P.H., et al. (2009). An oestrogen-receptor- α -bound human chromatin interactome. *Nature* 462, 58–64.
- Bailey, S.D., Zhang, X., Desai, K., Aid, M., Corradin, O., Cowper-Sal Lari, R., Akhtar-Zaidi, B., Scacheri, P.C., Haibe-Kains, B., and Lupien, M. (2015). ZNF143 provides sequence specificity to secure chromatin interactions at gene promoters. *Nat. Commun.* 2, 6186.
- Drissen, R., Palstra, R.J., Gillemans, N., Splinter, E., Grosveld, F., Philipsen, S., and de Laat, W. (2004). The active spatial organization of the beta-globin locus requires the transcription factor EKLF. *Genes Dev.* 18, 2485–2490.
- Vakoc, C.R., Letting, D.L., Gheldof, N., Sawado, T., Bender, M.A., Groudine, M., Weiss, M.J., Dekker, J., and Blobel, G.A. (2005). Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. *Mol. Cell* 17, 453–462.
- Splinter, E., Heath, H., Kooren, J., Palstra, R.J., Klous, P., Grosveld, F., Galjart, N., and de Laat, W. (2006). CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev.* 20, 2349–2354.
- Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S., et al. (2010). Mediator and cohesin connect gene expression and chromatin architecture. *Nature* 467, 430–435.
- Allen, B.L., and Taatjes, D.J. (2015). The Mediator complex: a central integrator of transcription. *Nat. Rev. Mol. Cell Biol.* 16, 155–166.
- Ji, X., Dadon, D.B., Abraham, B.J., Lee, T.I., Jaenisch, R., Bradner, J.E., and Young, R.A. (2015). Chromatin proteomic profiling reveals novel proteins associated with histone-marked genomic regions. *Proc. Natl. Acad. Sci. USA* 112, 3841–3846.

15. Zhang, K., Li, N., Ainsworth, R.I., and Wang, W. (2016). Systematic identification of protein combinations mediating chromatin looping. *Nat. Commun.* 7, 12249.
16. Duren, Z., Chen, X., Jiang, R., Wang, Y., and Wong, W.H. (2017). Modeling gene regulation from paired expression and chromatin accessibility data. *Proc. Natl. Acad. Sci. USA* 114, E4914–E4923.
17. Djekidel, M.N., Liang, Z., Wang, Q., Hu, Z., Li, G., Chen, Y., and Zhang, M.Q. (2015). 3CPET: finding co-factor complexes from ChIA-PET data using a hierarchical Dirichlet process. *Genome Biol.* 16, 288.
18. Cao, Q., Anyansi, C., Hu, X., Xu, L., Xiong, L., Tang, W., Mok, M.T.S., Cheng, C., Fan, X., Gerstein, M., et al. (2017). Reconstruction of enhancer-target networks in 935 samples of human primary cells, tissues and cell lines. *Nat. Genet.* 49, 1428–1436.
19. Whalen, S., Truty, R.M., and Pollard, K.S. (2016). Enhancer-promoter interactions are encoded by complex genomic signatures on looping chromatin. *Nat. Genet.* 48, 488–496.
20. Hait, T.A., Amar, D., Shamir, R., and Elkon, R. (2018). FOCS: a novel method for analyzing enhancer and gene activity patterns infers an extensive enhancer-promoter map. *Genome Biol.* 19, 56.
21. Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 376–380.
22. Gibcus, J.H., and Dekker, J. (2013). The hierarchy of the 3D genome. *Mol. Cell* 49, 773–782.
23. Gorkin, D.U., Leung, D., and Ren, B. (2014). The 3D genome in transcriptional regulation and pluripotency. *Cell Stem Cell* 14, 762–775.
24. Merkenschlager, M., and Nora, E.P. (2016). CTCF and Cohesin in Genome Folding and Transcriptional Gene Regulation. *Annu. Rev. Genomics Hum. Genet.* 17, 17–43.
25. Davidson, I.F., Bauer, B., Goetz, D., Tang, W., Wutz, G., and Peters, J.M. (2019). DNA loop extrusion by human cohesin. *Science* 366, 1338–1345.
26. Sanborn, A.L., Rao, S.S., Huang, S.C., Durand, N.C., Huntley, M.H., Jewett, A.I., Bochkov, I.D., Chinnappan, D., Cutkosky, A., Li, J., et al. (2015). Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci. USA* 112, E6456–E6465.
27. Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., and Mirny, L.A. (2016). Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep.* 15, 2038–2049.
28. Kim, Y., Shi, Z., Zhang, H., Finkelstein, I.J., and Yu, H. (2019). Human cohesin compacts DNA by loop extrusion. *Science* 366, 1345–1349.
29. Dang, C.V. (2012). MYC on the path to cancer. *Cell* 149, 22–35.
30. Miyashita, T., and Reed, J.C. (1993). Bcl-2 oncprotein blocks chemotherapy-induced apoptosis in a human leukemia cell line. *Blood* 81, 151–157.
31. Phillips, J.E., and Corces, V.G. (2009). CTCF: master weaver of the genome. *Cell* 137, 1194–1211.
32. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504.
33. Wang, Z., Cao, R., Taylor, K., Briley, A., Caldwell, C., and Cheng, J. (2013). The properties of genome conformation and spatial gene interaction and regulation networks of normal and malignant human cell types. *PLoS ONE* 8, e58793.
34. Ogryzko, V.V., Schiltz, R.L., Russanova, V., Howard, B.H., and Nakatani, Y. (1996). The transcriptional coactivators p300 and CBP are histone acetyltransferases. *Cell* 87, 953–959.
35. Visel, A., Blow, M.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., et al. (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* 457, 854–858.
36. Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* 39, 311–318.
37. Hennig, A.K., Peng, G.H., and Chen, S. (2013). Transcription coactivators p300 and CBP are necessary for photoreceptor-specific chromatin organization and gene expression. *PLoS ONE* 8, e69721.
38. Turnell, A.S., Stewart, G.S., Grand, R.J., Rookes, S.M., Martin, A., Yamano, H., Elledge, S.J., and Gallimore, P.H. (2005). The APC/C and CBP/p300 cooperate to regulate transcription and cell-cycle progression. *Nature* 438, 690–695.
39. Nesbit, C.E., Tersak, J.M., and Prochownik, E.V. (1999). MYC oncogenes and human neoplastic disease. *Oncogene* 18, 3004–3016.
40. Koh, C.M., Bezzi, M., Low, D.H., Ang, W.X., Teo, S.X., Gay, F.P., Al-Haddawi, M., Tan, S.Y., Osato, M., Sabò, A., et al. (2015). MYC regulates the core pre-mRNA splicing machinery as an essential step in lymphomagenesis. *Nature* 523, 96–100.
41. Ramos Pittol, J.M., Oruba, A., Mittler, G., Saccani, S., and van Essen, D. (2018). Zbtb7a is a transducer for the control of promoter accessibility by NF-kappa B and multiple other transcription factors. *PLoS Biol.* 16, e2004526.
42. Weichert, W., Boehm, M., Gekeler, V., Bahra, M., Langrehr, J., Neuhaus, P., Denkert, C., Imre, G., Weller, C., Hofmann, H.P., et al. (2007). High expression of RelA/p65 is associated with activation of nuclear factor-kappaB-dependent signaling in pancreatic cancer and marks a patient population with poor prognosis. *Br. J. Cancer* 97, 523–530.
43. Gayther, S.A., Batley, S.J., Linger, L., Bannister, A., Thorpe, K., Chin, S.F., Daigo, Y., Russell, P., Wilson, A., Soutter, H.M., et al. (2000). Mutations truncating the EP300 acetylase in human cancers. *Nat. Genet.* 24, 300–303.
44. Liu, L., Li, Q.Z., Jin, W., Lv, H., and Lin, H. (2019). Revealing Gene Function and Transcription Relationship by Reconstructing Gene-Level Chromatin Interaction. *Comput. Struct. Biotechnol. J.* 17, 195–205.
45. Cao, R., and Cheng, J. (2016). Integrated protein function prediction by mining function associations, sequences, and protein-protein and gene-gene interaction networks. *Methods* 93, 84–91.
46. Javierre, B.M., Burren, O.S., Wilder, S.P., Kreuzhuber, R., Hill, S.M., Sewitz, S., Cairns, J., Wingett, S.W., Varnai, C., Thiecke, M.J., et al. (2016). Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. *Cell* 167, 1369–1384.
47. Cairns, J., Freire-Pritchett, P., Wingett, S.W., Várnai, C., Dimond, A., Plagnol, V., Zerbino, D., Schoenfelder, S., Javierre, B.M., Osborne, C., et al. (2016). CHICAGO: robust detection of DNA looping interactions in Capture Hi-C data. *Genome Biol.* 17, 127.
48. Vian, L., Pekowska, A., Rao, S.S.P., Kieffer-Kwon, K.R., Jung, S., Baranello, L., Huang, S.C., El Khattabi, L., Dose, M., Pruett, N., et al. (2018). The Energetics and Physiological Impact of Cohesin Extrusion. *Cell* 173, 1165–1178.
49. Mathelier, A., Fornes, O., Arenillas, D.J., Chen, C.Y., Denay, G., Lee, J., Shi, W., Shyr, C., Tan, G., Worsley-Hunt, R., et al. (2016). JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* 44 (D1), D110–D115.
50. Wingender, E., Dietze, P., Karas, H., and Knüppel, R. (1996). TRANSFAC: a database on transcription factors and their DNA binding sites. *Nucleic Acids Res.* 24, 238–241.
51. Hume, M.A., Barrera, L.A., Gisselbrecht, S.S., and Bulyk, M.L. (2015). UniPROBE, update 2015: new tools and content for the online database of protein-binding microarray data on protein-DNA interactions. *Nucleic Acids Res.* 43, D117–D122.
52. Jolma, A., Yan, J., Whittington, T., Toivonen, J., Nitta, K.R., Rastas, P., Morgunova, E., Enge, M., Taipale, M., Wei, G., et al. (2013). DNA-binding specificities of human transcription factors. *Cell* 152, 327–339.
53. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* 38, 576–589.
54. Zerbino, D.R., Wilder, S.P., Johnson, N., Juettemann, T., and Flicek, P.R. (2015). The ensemble regulatory build. *Genome Biol.* 16, 56.
55. Chattri-Aryamontri, A., Oughtred, R., Boucher, L., Rust, J., Chang, C., Kolas, N.K., O'Donnell, L., Oster, S., Theesfeld, C., Sellam, A., et al. (2017). The BioGRID interaction database: 2017 update. *Nucleic Acids Res.* 45 (D1), D369–D379.
56. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842.