

RESEARCH ARTICLE

DLoopCaller: A deep learning approach for predicting genome-wide chromatin loops by integrating accessible chromatin landscapes

Siguo Wang¹, Qinhu Zhang^{2*}, Ying He¹, Zhen Cui¹, Zhenghao Guo¹, Kyungsook Han³, De-Shuang Huang^{4,5*}

1 Institute of Machine Learning and Systems Biology, School of Electronics and Information Engineering, Tongji University, Shanghai, China, **2** Translational Medical Center for Stem Cell Therapy and Institute for Regenerative Medicine, Shanghai East Hospital, Bioinformatics Department, School of Life Sciences and Technology, Tongji University, Shanghai, China, **3** The School of Computer Science and Engineering, Inha University, Incheon, South Korea, **4** EIT Institute for Advanced Study, Zhenhai, Ningbo, Zhejiang, China, **5** Big Data and Intelligent Computing Research Center, Guangxi Academy of Science, Nanning, China

* 20310031@tongji.edu.cn (QZ); dshuang@eias.ac.cn (DSH)



OPEN ACCESS

Citation: Wang S, Zhang Q, He Y, Cui Z, Guo Z, Han K, et al. (2022) DLoopCaller: A deep learning approach for predicting genome-wide chromatin loops by integrating accessible chromatin landscapes. *PLoS Comput Biol* 18(10): e1010572. <https://doi.org/10.1371/journal.pcbi.1010572>

Editor: Ferhat Ay, La Jolla Institute for Allergy and Immunology, UNITED STATES

Received: May 4, 2022

Accepted: September 14, 2022

Published: October 7, 2022

Copyright: © 2022 Wang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: These data are obtained from public datasets, including ENCODE (<https://www.encodeproject.org/>) with accession code ENCF264NMW, ENCF901GZH, ENCF013TGD, ENCF097SKJ, ENCF352SET, ENCF001THV and ENCF289WNN, NCBI (<https://www.ncbi.nlm.nih.gov/>) with accession code GSE137335, and 4DN portal (<https://data.4dnucleome.org/>) with accession code 4DNFIQNCHGRE, 4DNESR9S8R38 and 4DNFI6HDY7W. And the source codes of

Abstract

In recent years, major advances have been made in various chromosome conformation capture technologies to further satisfy the needs of researchers for high-quality, high-resolution contact interactions. Discriminating the loops from genome-wide contact interactions is crucial for dissecting three-dimensional(3D) genome structure and function. Here, we present a deep learning method to predict genome-wide chromatin loops, called DLoopCaller, by combining accessible chromatin landscapes and raw Hi-C contact maps. Some available orthogonal data ChIA-PET/HiChIP and Capture Hi-C were used to generate positive samples with a wider contact matrix which provides the possibility to find more potential genome-wide chromatin loops. The experimental results demonstrate that DLoopCaller effectively improves the accuracy of predicting genome-wide chromatin loops compared to the state-of-the-art method Peakachu. Moreover, compared to two of most popular loop callers, such as HiCCUPS and Fit-Hi-C, DLoopCaller identifies some unique interactions. We conclude that a combination of chromatin landscapes on the one-dimensional genome contributes to understanding the 3D genome organization, and the identified chromatin loops reveal cell-type specificity and transcription factor motif co-enrichment across different cell lines and species.

Author summary

The emergence of chromosome conformation capture technologies has provided researchers with the opportunity to understand the role of three-dimensional genome structure in regulating gene expression and cell functions. Although significant progress has been made in studying the basic functional units (called chromatin loops) that directly regulate gene expression, but still have limitations on how to adequately extract features from the contact maps and rationally utilize multi-omics data. In this work, we effectively

DLoopCaller are available at <https://github.com/wangguoqoa/DLoopCaller>.

Funding: This work was supported by the grant of National Key R&D Program of China (No. 2018YFA0902600 & 2018AAA0100100) and partly supported by National Natural Science Foundation of China (Grant nos. 62002266, 61932008, and 62073231), and Introduction Plan of High-end Foreign Experts (Grant no. G2021033002L) and, respectively, supported by the Key Project of Science and Technology of Guangxi (Grant no. 2021AB20147), Guangxi Natural Science Foundation (Grant nos. 2021JJA170204 & 2021JJA170199) and Guangxi Science and Technology Base and Talents Special Project (Grant nos. 2021AC19354 & 2021AC19394) to DSH. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

combine accessible chromatin landscapes and raw Hi-C contact maps data based on a deep learning framework to identify genome-wide chromatin loops. Besides, we use some available orthogonal data ChIA-PET/HiChIP and Capture Hi-C were used to generate training samples. We demonstrate the performance of our proposed method to identify some unique chromatin loops with high confidence. Moreover, the identified chromatin loops further reveal cell-type specificity and transcription factor motif co-enrichment across different cell lines and species, which may help us understand the mechanism of tissue-specific gene expression and transcriptional regulation.

This is a *PLOS Computational Biology Methods* paper.

Introduction

In eukaryotes, chromatin is folded into complex 3D structures and dynamically regulates the life processes. Therefore, dissecting the rules that govern chromatin dynamics is essential to comprehend the tissue-specific gene regulation, which provides the rationale for understanding the role of noncoding region variants associated with disease [1–3]. In the past two decades, many high-throughput technologies have emerged for researchers to reveal the significance of chromatin structure for gene regulatory networks. According to these technologies, from a genome-wide perspective, the multiscale high-dimensional chromatin structure is divided into A/B compartments, more refined nuclear compartmentalization, topologically associating domains (TADs), and chromatin loops [4–9]. Gene regulatory networks rely on cis-regulatory elements, for example, many enhancers function over long genomic distances to regulate gene expression by forming topological loops with distant promoters and form an active chromatin hub consisting of multiple enhancers and their interacting promoters [10–12]. In addition, ChIA-PET data indicate that architectural proteins play an important role in forming chromatin structure and regulating transcription, including CCCTC-binding factor (CTCF), cohesin, and RNA polymerase II [13]. Sanborn et al. revealed that chromatin loops are mediated by two pairs of structural proteins CTCF and cohesin in a loop extrusion model, and until the corresponding CTCF is detected on the strand will stop [14]. The spatial chromatin structure is not only characterized by gene expression but also conserved across species [15]. Although several studies have given significant insights into 3D genome organization and function, they still lack of capacity to describe chromatin loops in the 3D space of the nucleus and predict the impact of structural changes on genetic mutations. Understanding the relationship between the complex structure and function of the genome remains a big challenge, hence more computational models are urgently needed to be proposed for 3D genomic studies.

Numerous experimental methods have been developed to predict 3D chromatin loops, mainly divided into the following aspects: (1) Sequencing-based techniques. High-throughput chromosome conformation capture (Hi-C) aims to sequence 3D interactions at the genome-wide level, which include dilution during proximity ligation but is less effective [7,8]. The ensuing in situ ligation compensates for the deficiencies, efficiently capturing true contacts and providing higher resolution at the same sequencing depth. GAM and SPRITE are used to analyze two-way and multi-way contacts, enabling the direct study of multivalent enhancer-promoter interactions [16,17]. However, to finely map chromatin folding and understand

some of its functional aspects, it is necessary to detect specific contacts using enrichment methods that amplify the contact signal in specific genomic regions of interest. Other 3C-based technologies have been proposed, Capture-HiC technology captures chromatin interaction maps in specific regions (such as promoter regions) through hybridization probes, which is low-cost but achieves deeper-depth sequencing [18]. And some interaction maps are mediated by the specific proteins of interest, such as Chromatin Interaction Analysis with Paired-End Tag sequencing (ChIA-PET) [19] and HiChIP [20]. (2) Super-resolution microscopy methods. Stochastic optical reconstruction microscopy (STORM) [21] and Structured illumination microscopy (SIM) [22] are two classical methods to illustrate the power of single-cell super-resolution imaging.

With the advent of Hi-C and related technologies, some computational analysis tools, Fit-Hi-C [23] and HiCCUPS [8,24] are the two most popular enrichment-based methods, have been proposed to study the inherent complexity of Hi-C data. Fit-Hi-C model the random polymer looping effect to assign statistical confidence with genomic distance into account, specific chromatin contacts remarkably increase about contact detection compared with general background model. HiCCUPS identifies “enriched pixels” as chromatin loops which means comparing the number of contacts in the pixel with a series of regions surrounding the pixel. Although these computational tools have made great progress, they still have some drawbacks, such as high cost and conservative. All of these limitations have stimulated the development of computational analyses and mathematical models, combined with experimental methods, which may quantitatively and predictively understand chromosome structure and function. For example, CHiCAGO applies a convolution background model to predict DNA looping interactions in Capture Hi-C data [25]. To date, there are some studies to predict CTCF-mediated chromatin interactions based on a random forest model by integrating genomic, epigenomic features, or transcription factor profiles [26,27]. Owing to the rapid development and widespread application of deep learning techniques, it is not surprising that significant progress has been made in bioinformatics [28–32], and some works have been made in the field of genomics. For instance, Deep-loop predicted CTCF-mediated chromatin loops and performs well in different cell lines [33], and DeepMILO predicted the effects of variants on CTCF and cohesin-mediated insulator loops based on a deep learning framework [34]. Furthermore, Mustache employed scale-space theory in computer vision to detect chromatin loops in contact maps, regarding only locally enriched pixels as loops [35]. It is worth noting that Peakachu built a random forest to predict chromatin loops in genome-wide contact maps, which transforms the task of detecting chromatin loops into a binary classification problem by using biologically enriched experiments such as ChIA-PET/HiChIP and Capture Hi-C as positive samples and non-interaction regions as negative samples, achieving impressive prediction performance [36].

Although these methods have achieved breakthroughs, how to adequately extract features from the contact maps and rationally utilize multi-omics data to identify chromatin loops is still a big challenge. In this study, we present a new method named DLoopCaller, based on a deep learning framework, for predicting chromatin loops in genome-wide contact maps by integrating raw Hi-C matrix and accessible chromatin landscape. Similar to Peakachu, DLoopCaller transforms the task of detecting chromatin loops into a binary classification problem by using enriched experimental data such as ChIA-PET/HiChIP and Capture Hi-C as positive interactions and non-interaction regions as negative samples. The contributions of DLoopCaller mainly include the following aspects: (i) efficiently combining one dimensional (1D) open chromatin landscapes with 3D genomic data for chromatin loops prediction; (ii) improving the identification accuracy of chromatin loops on wider chromatin contact matrix; (iii) and compared with some existing methods, our method identifies a series of unique chromatin

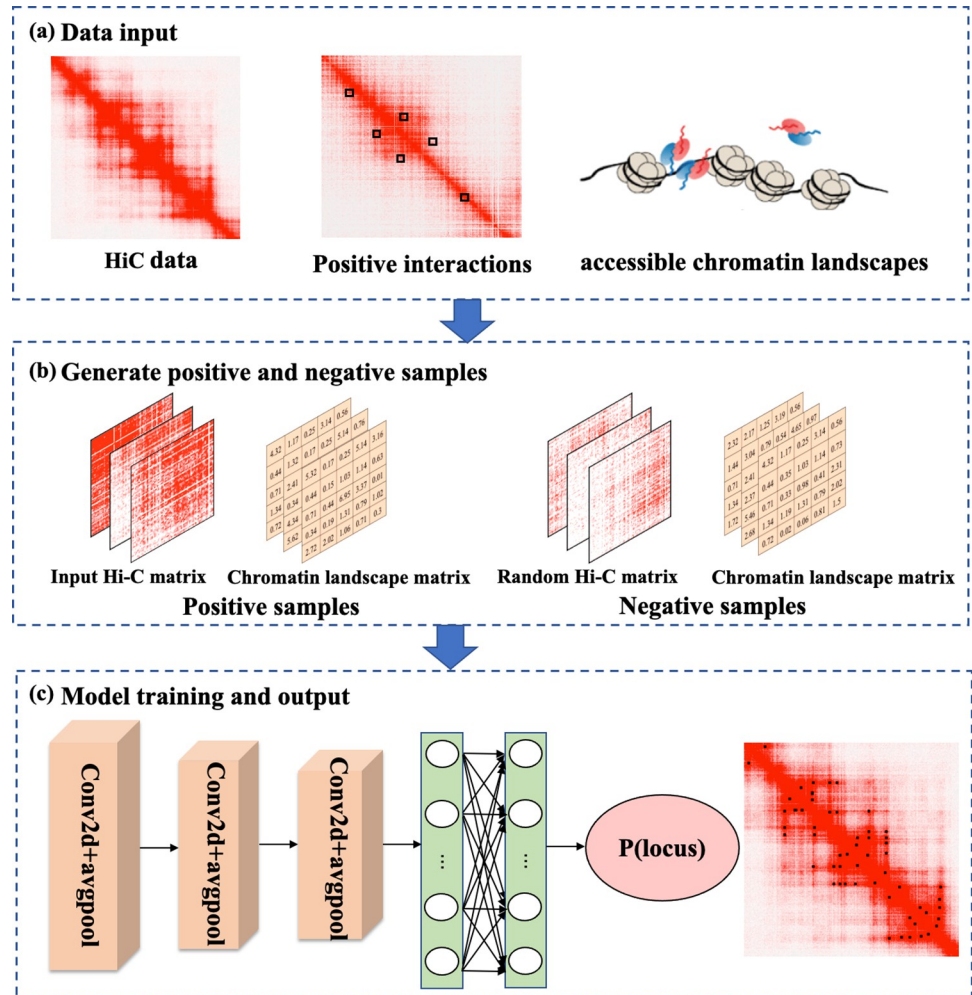


Fig 1. The workflow of DLoopCaller. (a) Data inputs includes Hi-C matrix, accessible chromatin landscapes, and enriched experimental data such as ChIA-PET/HiChIP and Capture Hi-C as positive interactions. (b) Positive samples are generated according to the input data, and negative samples are generated according to the similar distance or greater distance of the positive samples. (c) DLoopCaller includes three convolutional blocks, two fully connected layers and a classification layer, in which each block consists of a convolutional layer, a ReLU layer, a dropout layer, and followed by a global average pooling layer.

<https://doi.org/10.1371/journal.pcbi.1010572.g001>

loops at 10 kb in genome-wide contact maps; (iv) the identified chromatin loops reveal cell-type specificity and transcription factor motif co-enrichment; (v) DLoopCaller is robust and reproducible to some extent. The workflow of DLoopCaller is shown in Fig 1.

Materials and methods

Data collection and preprocessing

We performed experiments on four cell lines, including K562 (chronic myelogenous leukemia), GM12878 (lymphoblastoid cell), H1-ESC (hematopoietic stem cell), and mESC (mouse embryonic stem cells), and the data inputs include Hi-C data, accessible chromatin data, and corresponding enriched experimental data. The original Hi-C data were converted into 10kb resolution contact matrices and normalized by using hic2cool and cooler python package.

The Hi-C contact maps of GM12878 can be downloaded from <https://drive.google.com/file/d/1rfkdHSfmn5GK7qdzSwVlrSHpJVPPn5R3/view?usp=sharing>. In order to reduce data bias, we merged the accessible chromatin landscapes of two replicate samples as the final data of GM12878, which were obtained from ENCODE with accession code ENCFF264NMW and ENCFF901GZH. The enriched experimental data in GM12878 include CTCF ChIA-PET interactions [13], Rad21 ChIA-PET interactions [37], SMC1 HiChIP interactions [20], H3K27ac HiChIP interactions [1] and promoter Capture Hi-C interactions [25].

The Hi-C contact maps of K562 were obtained from the ENCODE with accession code ENCFF013TGD (replicate1) and ENCFF097SKJ (replicate2). The accessible chromatin landscape of K562 was obtained from ENCODE with accession code ENCFF352SET. The enriched experimental data CTCF ChIA-PET interactions in K562 were obtained from ENCODE with accession code ENCFF001THV.

The Hi-C contact maps of H1-ESC were obtained from the 4DN data portal with accession code 4DNFI6HDY7W. The accessible chromatin landscape of H1-ESC was obtained from the 4DN data portal with accession code 4DNFIQNCHGRE. The enriched experimental data CTCF ChIA-PET interactions in H1-ESC were obtained from the 4DN data portal with accession code 4DNESR9S8R38.

The Hi-C contact maps of mESC were obtained from ENCODE with accession code ENCFF289WNN. The accessible chromatin landscape of the mESC was obtained from NCBI with accession code GSE137335. The enriched experimental data SMC1 HiChIP interactions in mESC were obtained from [20].

All mentioned positive interactions obtained from enrichment experiments are consistent with Peakachu, provided at <https://github.com/wanguoguo/DLoopCaller/tree/main/training-sets>. The enhancer and promoter loci in GM12878, K562, H1-ESC provided at <https://github.com/wanguoguo/DLoopCaller/tree/main/annotations>, which were extracted from public ChromHMM annotations in ENCODE.

Methods

The generation of training samples

The data inputs of DLoopCaller mainly include three parts: the original Hi-C matrix, some verified positive interactions involving targeted regions or proteins of interest by biologically enriched experiments such as ChIA-PET/HiChIP and Capture Hi-C, and the corresponding accessible chromatin landscapes, which were then used to generate training samples for training model. Briefly, (i) The pixels around each positive interaction were used as the features of the training samples, in which the pixel of the positive interaction was expanded along both sides in the raw HiC matrix to obtain a 23×23 positive Hi-C matrix; (ii) In order to obtain the corresponding accessible chromatin matrix, the 1D accessible chromatin data were firstly averaged at every 10kb distance to keep the resolution consistency and reduce the data deviation. Then the chromatin accessible data of the x -axis peak loci and y -axis peak loci in the positive HiC matrix were used to obtain the positive accessible chromatin matrix by Cartesian product. For example, $x = \{X_1, X_2, \dots, X_n\}$ and $y = \{X_1, X_2, \dots, X_n\}$, the accessible chromatin matrix is defined as follows (Fig 2):

where $n = 23$ and the blue matrix is the accessible chromatin matrix; (iii) the negative Hi-C matrix with an equal number of pixels from nonzero values was randomly sampled from two aspects: (1) matching the similar distance of positive interactions according to the probability density function of the distance; (2) considering greater distance, larger than maximum distance of the positive interactions, to improve the diversity of negative samples. Similarly, we obtained the corresponding accessible chromatin matrix for the negative HiC matrix following the same way described above. And we list the number of samples in each dataset in S1 Table.

	X_1	X_2	...	X_{n-1}	X_n
Y_1	X_1Y_1	X_2Y_1	...	$X_{n-1}Y_1$	X_nY_1
Y_2	X_1Y_2	X_2Y_2	...	$X_{n-1}Y_1$	X_nY_1
...
Y_{n-1}	X_1Y_{n-1}	X_2Y_{n-1}	...	$X_{n-1}Y_1$	$X_{n-1}Y_1$
Y_n	X_1Y_n	X_2Y_n	...	$X_{n-1}Y_1$	$X_{n-1}Y_1$

Fig 2. The generation of accessible chromatin matrix.

<https://doi.org/10.1371/journal.pcbi.1010572.g002>

The framework of neural network architecture

Some studies have shown that three-layer convolutional neural networks (CNN) are sufficient to mine features from complex biological data to achieve good experimental results [38–41]. Therefore, DLoopCaller applied a three-layer CNN model to extract features from the generated Hi-C matrix and accessible chromatin landscape matrix and retained the best training model for identifying genome-wide chromatin loops. As shown in Fig 1, DLoopCaller takes two-channel as input into the model, inspired by the way of image processing research. To reduce data bias and noise, the positive/ negative Hi-C matrix and accessible chromatin landscape matrix were normalized before training. The normalization of each matrix is as follows:

$$M(x, y) \leftarrow \log_{10}(1 + M(x, y))$$

$$M(x, y) \leftarrow \frac{M(x, y)}{1 + \max(M(x, y))} \tag{1}$$

where x, y refers to the coordinates of the Hi-C matrix or the accessible chromatin landscape matrix M , respectively, $\max(M(x, y))$ denotes the maximum value in the corresponding positive/negative Hi-C matrix and accessible chromatin landscape matrix.

The framework of DLoopCaller is composed of three convolutional blocks, in which each block consists of a convolutional layer, a ReLU layer, a dropout layer, and a global average pooling layer. The convolutional layer is used to directly capture local features and the global

average pooling layer is used to capture the global textual information from the Hi-C matrix and chromatin landscape matrix, and the dropout layer is used to avoid falling into overfitting and reduce complex co-adaptation relationships between neurons, which is set to 0.2. And then two fully connected layers of 64 neurons are used to fuse the features of the Hi-C matrix and chromatin landscape matrix. Meanwhile, the batch-normalization layer is used to speed up the model convergence and prevent the gradient explosion and disappearance during the calculation process. Finally, the sigmoid layer is used to output the probabilities of candidate chromatin loops. A more detailed description of the framework is shown in [S2 Table](#).

Model training

For a fair comparison with the competing methods, we used the leave-one-out for training, validation, and testing. More specifically, 22 chromosomes were used for training and validation, where 80% of chromosomes is used for training and 20% of chromosomes is used for validation, and the remaining one chromosome is used for testing. DLoopCaller regards the prediction of genome-wide chromatin loops as a binary classification task, hence the binary cross-entropy loss (BCELoss) is used for training model, which is defined as follows:

$$BCE = -Y_i * \log(\bar{Y}_i) - (1 - Y_i) * \log(1 - \bar{Y}_i) \quad (2)$$

where Y_i denotes original values and \bar{Y}_i denotes the predicted value of the i -th sample. The BCELoss is optimized by the Adam optimization algorithm [42] with a batch size is 128, and the learning rate is set to 0.001. During the training, DLoopCaller applied five-fold cross-validation to iteratively select the best parameters for distinguishing whether it is a chromatin loop and saved the model. Our proposed model is written by Python based on the Pytorch framework. We used a machine with Tesla K40 GPU with 10GB memory for training on the Linux system.

Identifying genome-wide chromatin loops

Once the best model for each chromosome is trained, it can be used to predict all potential chromatin loops in the corresponding chromosome. Identifying chromatin loops from the whole genome includes two stages: one is to use the best trained model to score all potential chromatin loops, and the other is to pool candidate chromatin loops. Firstly, we used the best trained model to score all non-zero pixels meaning potential chromatin loops on each chromosome. Some studies have shown that the higher the interaction frequency in the Hi-C map, the greater the probability of becoming a chromatin loop [8]. Hence, to accurately and efficiently predict chromatin loops, DLoopCaller only retained those candidate chromatin loops whose contact frequency is greater than the average of all candidate chromatin loops. Finally, we used the greedy algorithm provided by Peakachu [36] to cluster all candidate chromatin loops and selected the most representative pixels as the identified chromatin loops.

Results

The overall performance of DLoopCaller on different cell lines

To train and measure the performance of our proposed method DLoopCaller, we performed DLoopCaller on three human cell lines (GM12878, K562 (replicate1), H1-ESC) and a mouse cell line (mESC) to validate its classification performance. The F1-score and PRAUC (Area Under the Precision-Recall Curve) metrics were employed to verify the performance of DLoopCaller and competing methods in distinguishing whether it is a chromatin loop, which were defined in [S1 Note](#). Since the recent methods are limited to identifying a protein of

interest-mediated chromatin loops [27,33], we mainly compared the proposed method DLoopCaller with a comprehensive method Peakachu in this part, and all the same enriched data were performed on both methods separately to validate the effectiveness of deep learning framework and accessible chromatin landscapes. Peakachu used the interaction frequency and rank as features in smaller matrices based on a random forest approach, outperforming Gaussian Naïve Bayes, Perceptron, Logistic Regression, SVM (linear kernel), and SVM (RBF kernel). In the GM12878 cell line, five enriched experimental data were used to label positive samples and train the model, including CTCF ChIA-PET, H3K27ac HiChIP, SMC1 HiChIP, RAD21 ChIA-PET, and promoter Capture Hi-C. The corresponding CTCF ChIA-PET in K562 (replicate1) and H1-ESC, and SMC1 HiChIP in mESC were separately used to train the model. To comprehensively evaluate the classification performance of DLoopCaller, we used the average value of F1-score, PRAUC, Precision and Recall for all chromosomes in each cell line.

Our proposed method DLoopCaller uses deep learning framework to automatically learn features instead of hand-designed features used in Peakachu for the identification of chromatin loops, which is one of the innovations of our approach. In order to fully extract the features of chromatin loops, we use a larger window (23×23) to generate the feature matrix of positive and negative samples. To better illustrate this issue, we extend DLoopCaller with window 11×11 on H3K27ac HiChIP and RAD21 ChIA-PET GM12878. As shown in Fig 3(A), we can see that even though DLoopCaller uses 11×11 window size of each interaction to generate features, the performance of it is better than peakachu overall. The experimental results confirm our assumption that DLoopCaller with a larger window (23×23) performs better than Peakachu to identify chromatin loops. The window size is only a parameter, where larger window better fit DLoopCaller to improve the identification accuracy of chromatin loops. Although Peakachu and DLoopCaller use different window size, it is relatively fair for experimental purposes. And another innovation of our method is to efficiently combine one dimensional (1D) open chromatin landscapes with 3D genomic data for chromatin loops prediction, which is also not considered by Peakachu. Therefore, a larger window size (23×23) is adopted in DLoopCaller for the following experiments.

From Fig 3, we can see that the average of both F1-score and PRAUC of DLoopCaller are greater than Peakachu, indicating that the classification performance of DLoopCaller is better than that of Peakachu on all cell lines. As shown in Fig 3(B) and 3(C), which shows the experimental results of five enriched datasets in GM12878, we can see that the average F1-score and PRAUC value in the CTCF ChIA-PET, H3K27ac HiChIP, SMC1 HiChIP, RAD21 ChIA-PET are both close to 0.95 or greater than 0.95, showing a relatively excellent classification performance. It is worth noting that the F1-score of method DLoopCaller is about 8% higher than Peakachu in K562 (replicate1), but the F1-score and PRAUC value of the two methods are relatively lower compared to other cell lines. And the line and box plots of detailed results about F1-score and PRAUC were shown in S1 and S2 Figs, the performance of DLoopCaller obviously outperforms Peakachu on most of chromosomes. According to the boxplots of precision and recall values shown in S3 Fig, DLoopCaller is better than Peakachu except the precision of DLoopCaller is slightly lower than that of Peakachu in H1-ESC. The overall experimental results show that DLoopCaller combining Hi-C contact maps with accessible chromatin data to facilitate the prediction of genome-wide chromatin loops.

Performance assessment from different enriched experimental data within individual cell types

In order to further assess the performance of the proposed DLoopCaller, chromatin loops predicted from genome-wide contact maps were analyzed. We firstly performed experiments on

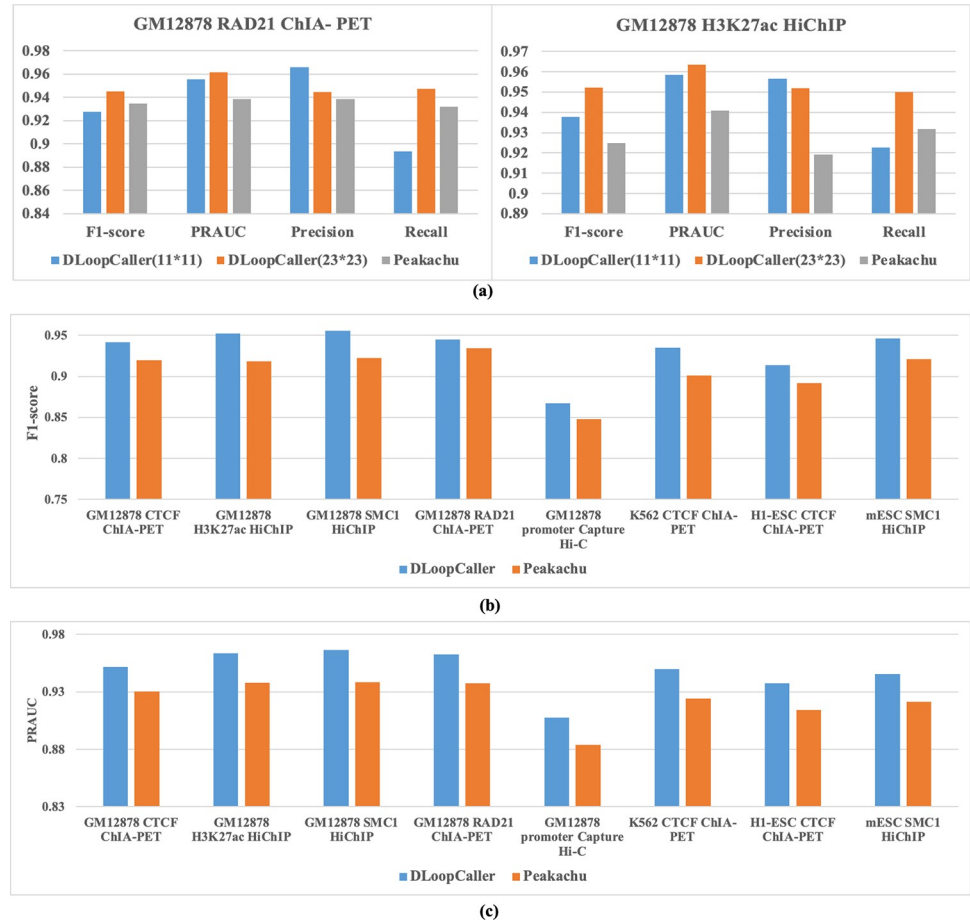


Fig 3. The overall performance comparison of DLoopCaller and Peakachu on different cell lines. (a) The F1-score, PRAUC, Precision and Recall values of RAD ChIA-PET and H3K27ac HiChIP in GM12878. (b)-(c): The F1-score and PRAUC values of CTCF ChIA-PET, H3K27ac HiChIP, SMC1 HiChIP, RAD ChIA-PET, and promoter Capture Hi-C in GM12878, CTCF ChIA-PET in K562 (replicate1) and H1-ESC, and SMC1 HiChIP in mESC.

<https://doi.org/10.1371/journal.pcbi.1010572.g003>

CTCF ChIA-PET, H3K27ac HiChIP, SMC1 HiChIP, RAD21 ChIA-PET, and promoter Capture Hi-C in GM12878, and analyzed the differences of predicted chromatin loops within individual cell types. The best trained model of each chromosome was used to predict chromatin loops, and then we aggregated the identified chromatin loops on all chromosomes for further analysis. As shown in Fig 4(A), the distance distribution of the identified chromatin loops in different enrichment experimental data varies, for example, the distance distributions of SMC1 HiChIP, RAD21 ChIA-PET, and promoter Capture Hi-C is similar, mainly located in 250kb and 500kb, and the proportions are 46.3% (7365/15908), 48.8%(9560/19590) and 49.1% (7097/14460). While CTCF ChIA-PET and H3K27ac HiChIP loops are mainly located in less than 250kb, the latter is about 5% higher than the former. The results confirm that the distances of long-range interactions are correlated with the factor of interest when using the same sequencing method [43].

To further assess this difference, the Aggregated Peak Analysis (APA) was used to quantify how well each chromatin loop set was supported by the Hi-C signals [7]. The APA plots of chromatin loops captured by the five enriched experiments in GM12878 are shown in S4 Fig. These APA plots show considerable enrichment compared to their local background and show strong consistency in GM12878 using different enrichment experiments. As shown in Fig 4

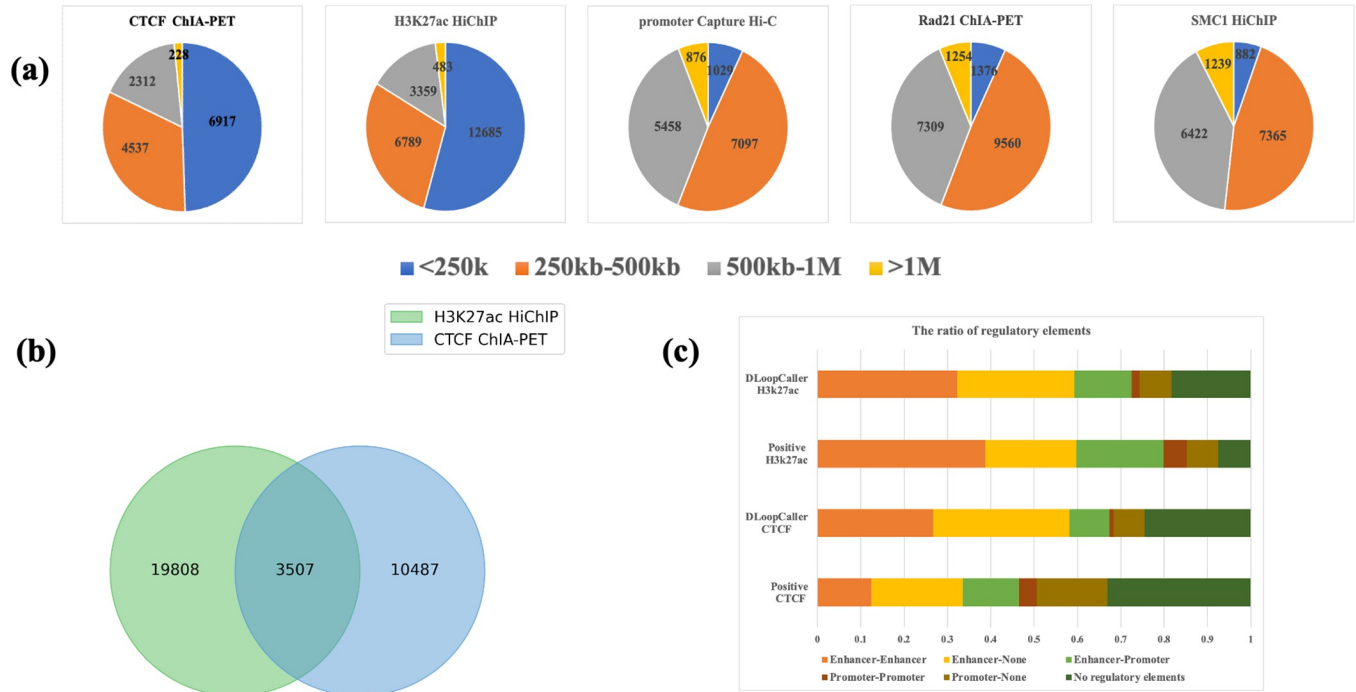


Fig 4. Comparison of chromatin loops within individual cell types. (a) Distance distribution of DLoopCaller identified chromatin loops from Hi-C contact maps by using CTCF ChIA-PET, H3K27ac HiChIP, SMC1 HiChIP, RAD ChIA-PET, and promoter Capture Hi-C data after training on GM12878. (b) Venn diagram of DLoopCaller identified chromatin loops determined by CTCF ChIA-PET and H3K27ac HiChIP experiments in GM12878. (c) The proportion of CTCF ChIA-PET interactions and H3K27ac HiChIP interactions types for GM12878. The proportion of identified chromatin loops types using CTCF ChIA-PET data and H3K27ac HiChIP after training for GM12878.

<https://doi.org/10.1371/journal.pcbi.1010572.g004>

(B), the overlapping chromatin loops of CTCF ChIA-PET and H3K27ac HiChIP loops only account for a quarter, which means the two anchors completely matched of two bins, even though the distance distribution of both is similar.

Some studies have shown that H3K27ac is an active enhancer- and promoter-associated histone marker and H3K27ac HiChIP can identify functional enhancer-promoter interactions with high confidence [1,44], and CTCF ChIA-PET aims to detect the specific long-range interactions [43]. Therefore, we analyzed the proportion of regulatory elements in the H3K27ac HiChIP data and identified H3K27ac HiChIP chromatin loops in GM12878. We find that the majority of the interactions and the identified loops in H3K27ac HiChIP data are mediated by enhancers, and the ratios are very close accounting for about 80% and 75% respectively. Compared to H3K27ac HiChIP data, the interactions in CTCF ChIA-PET are relatively smaller accounting for 47%, but the interactions without regulatory elements are relatively larger accounting for 30%. And the majority of identified chromatin loops in CTCF ChIA-PET are enhancer-mediated but have more long-range interactions. These results suggest that DLoopCaller is able to predict enhancer-regulated chromatin loops with high sensitivity, which may contribute to deciphering the principles of gene expression and disease-associated genetic variants.

Comparison of chromatin loops identified by different methods

To further validate the performance of the proposed DLoopCaller and increase the confidence of the identified chromatin loops, we compared CTCF ChIA-PET loops identified by DLoopCaller with some of the most popular methods, including Peakachu, global enrichment-based

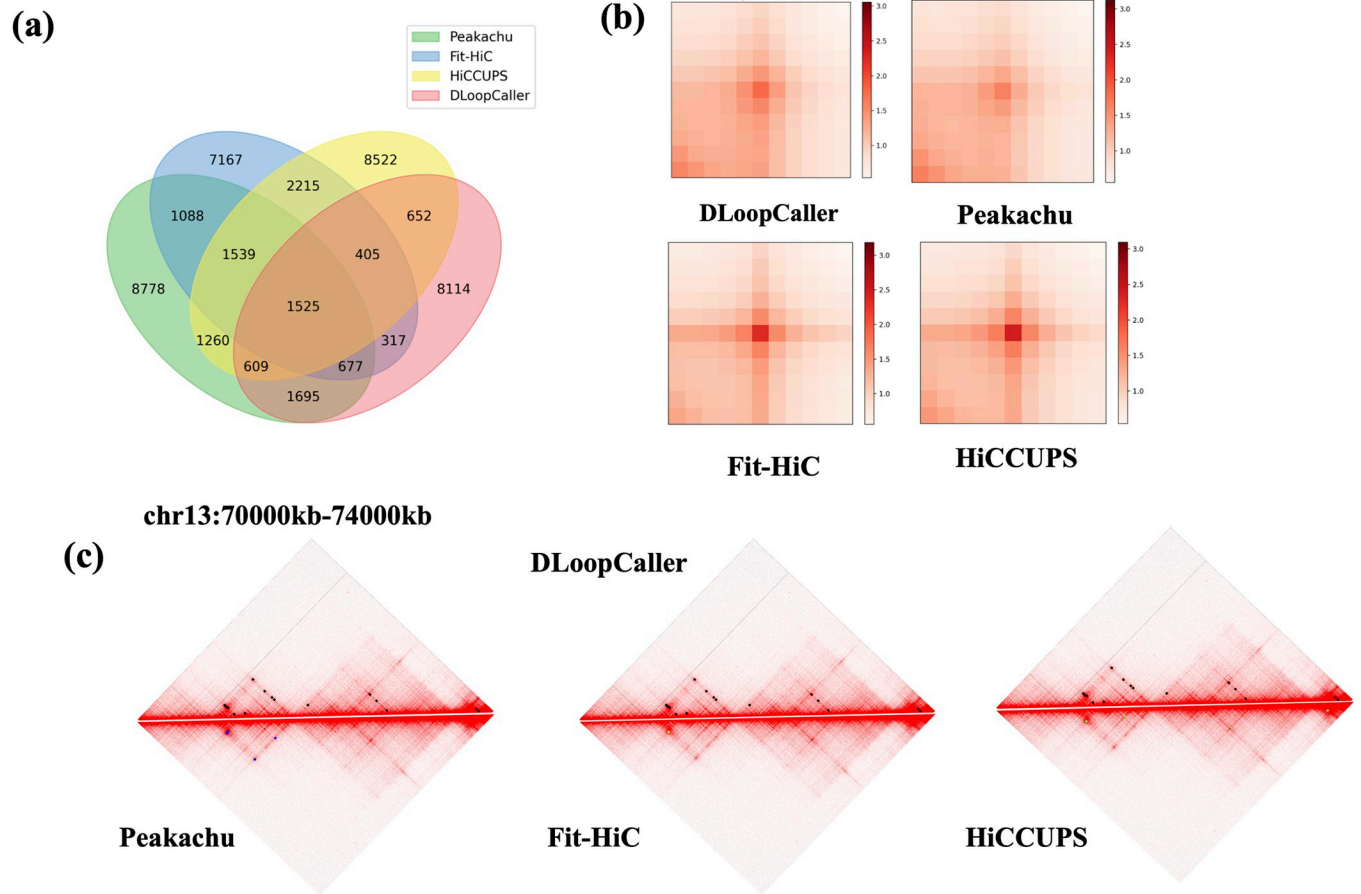


Fig 5. Comparison of chromatin loops identified by DLoopCaller, HiCCUPS, Fit-HiC, and Peakachu in GM12878. (a) Venn diagram of identified chromatin loops determined by DLoopCaller CTCF ChIA-PET, HiCCUPS, Fit-HiC, and Peakachu in GM12878. (b) APA plots for DLoopCaller CTCF ChIA-PET loops, HiCCUPS, Fit-HiC, and Peakachu in GM12878. (c) A visual example of identified loops by different models in a region. The black dots in the upper half of the three diamond-shaped graphs represent the chromatin loops identified by DLoopCaller, and the blue, green, and yellow dots in the lower half represent the chromatin loops identified by Peakachu, Fit-HiC, and HiCCUPS respectively.

<https://doi.org/10.1371/journal.pcbi.1010572.g005>

method Fit-Hi-C, and local enrichment-based methods HiCCUPS. For a fair comparison, the competing methods were also performed at 10kb resolution in GM12878 respectively and the identified chromatin loops were filtered to maintain the close number of DLoopCaller. We first compared the chromatin loops identified by each method by considering overlapping when the anchors of the two chromatin loops matched completely. As shown in Fig 5(A), we find that 42% (5880/13994) of identified chromatin loops by DLoopCaller are overlapped with ones by the other three methods, and 8114 chromatin loops are unique. We specifically compared CTCF ChIA-PET chromatin loops and H3K27ac HiChIP chromatin loops respectively identified by DLoopCaller and Peakachu. As shown in S5(A) and S5(B) Fig, the number of CTCF ChIA-PET loops of Peakachu and DLoopCaller is basically the same with the overlapping ratio 29.3%(4105/13994), and DLoopCaller identifies more H3K27ac HiChIP loops than Peakachu with the overlapping ratio 18.17% (4236/23315). From the perspective of the distance distribution of the identified chromatin loops, the distance distributions of the H3k27ac HiChIP loops identified by the two methods are very similar, but the proportion of long-range (>250kb) CTCF ChIA-PET chromatin loops identified by DLoopCaller is slightly lower than that of Peakachu (S5(C) and S5(D) Fig). The APA plots are used to inspect the overall loop patterns of the detected peaks by all methods, the APA plots of Fit-HiC and HiCCUPS show

strong consistency mainly focusing on the center pixel. Overall, the APA plots of DLoopCaller and Peakachu show similar enrichment of contact signals compared to surrounding pixels, but the former has a slightly stronger enrichment signal concentrating on the center pixel than the latter.

Taken together, the genome-wide analysis described above demonstrates that DLoopCaller has a good capability in terms of the identified loops from HiC contact maps. To further illustrate this point, we used juicebox (<https://github.com/aidenlab/Juicebox/wiki>), a visualization tool embedded in the juicer tool [24], to visualize some examples of the identified loops. We can see from Fig 5(C) that most of the chromatin loops identified by DLoopCaller and other methods in this region are overlapped but some are unique. And more visual examples are shown in S6 Fig.

Chromatin loops reveal cell-type specificity

Next, we evaluated the ability to identify loops in other cell lines. DLoopCaller separately identified 13994 CTCF ChIA-PET loops in GM12878, 10767 SMC1 HiChIP loops in H1-ESC, and 11841 CTCF ChIA-PET loops in K562 (replicate1), of which the short-range (< 250kb) interactions account for 49.4%, 87.2%, and 88.8%, respectively. To further illustrate the differences, we analyzed the APA profiles of all identified chromatin loops in three cell lines as shown in Fig 6(B). We find that the most important predictor is the center pixel and bottom left pixel respectively in GM12878 and H1-ESC, while it is jointly driven by the center and bottom left pixel in K562 (replicate1). In addition, we compared the overlap of chromatin loops in the three cell lines, and any anchors in the two bins are allowed to be incompletely matched to increase the fault tolerance. Briefly, two chromatin loops were considered matched if the ± 10 kb region around the center of one loop overlaps another. The comparison results are shown in Fig 6(C), even the tolerance for overlapping is increased, the overlapping chromatin loops of the three cell lines are also relatively less, which proves that the identified chromatin loops are cell-type specific.

In addition, to further analyze the relationship between the chromatin loops identified by DLoopCaller and the regulatory elements in all cell lines. As shown in Fig 6(D), the proportion of each regulatory element of SMC1 HiChIP data and identified chromatin loops in H1-ESC is very similar, which indicates that the chromatin loops identified by the DLoopCaller are reliable and demonstrates that the proposed DLoopCaller is effective. From Figs 4(C) and 6(D), we can conclude that the ratio of the regulatory elements in the chromatin loops identified by DLoopCaller and training data is basically the similar, and most of chromatin loops are regulated by enhancers. This experimental result has also been verified by the existing research [45], which provides the possibility to further understand the gene regulatory network. We also found that these predicted cell-type-specific loops are often located chromatin open regions and active enhancer regions (S7 Fig).

Transcription factor motif co-enrichment across different cell lines and species

Some studies have demonstrated that enhancer-promoter interactions regulate target genes in the genome, and specific transcription factor cooperation offers the possibility to understand the cell-type specificity of genome interactions [46], of which Cicero, PEP, and the latest proposed Spatzie attempted to detect the transcription factor motif cooperativity between enhancer-promoter interactions [47–49]. To analyze whether the sequence-based features within identified chromatin loops, we first performed experiments using Spatzie where all identified chromatin loops were used in each cell line. We applied Spatzie with count

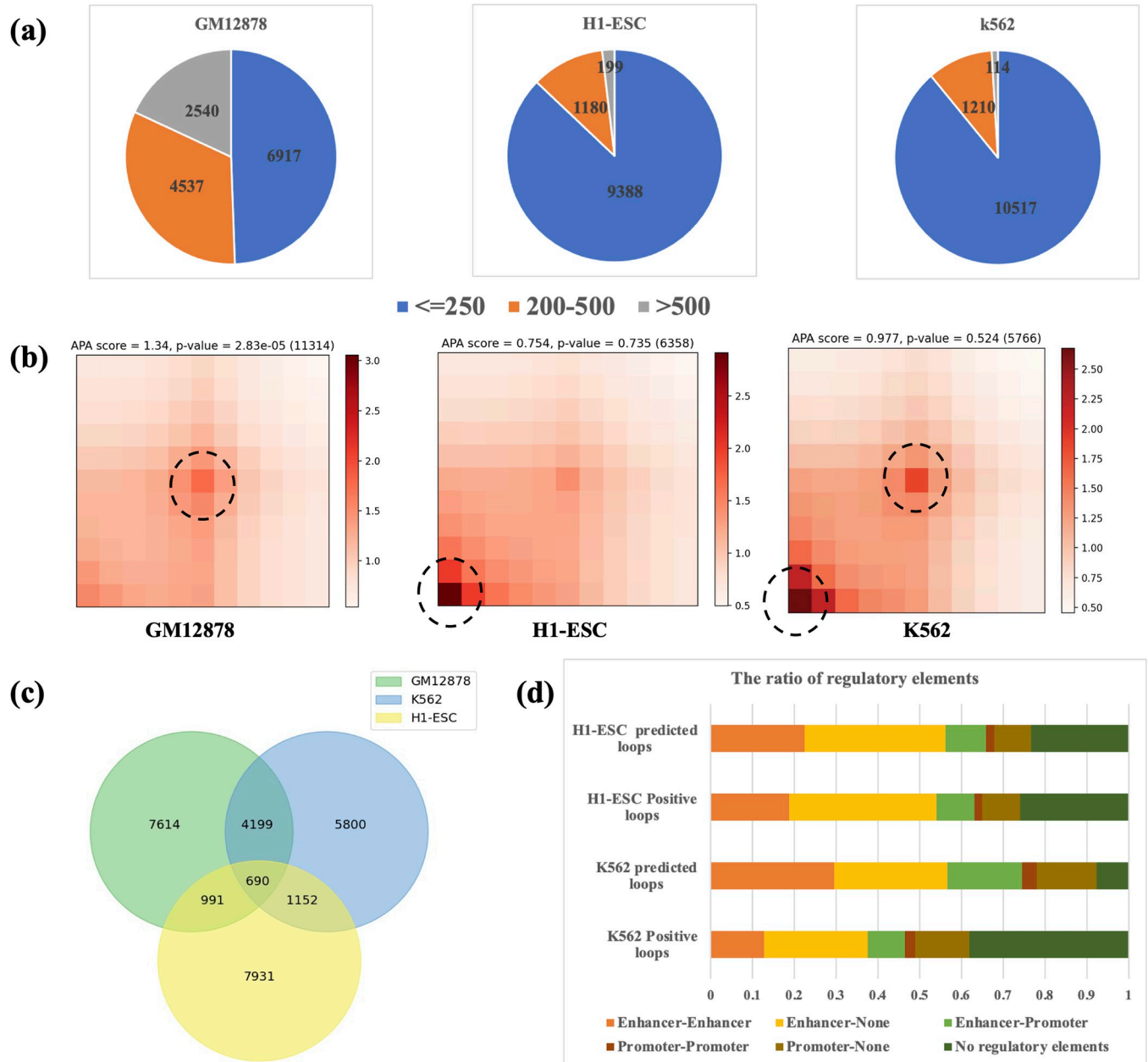


Fig 6. Comparison of chromatin loops among cell types. (a) Distance distribution of DLoopCaller identified chromatin loops from Hi-C contact maps by using CTCF ChIA-PET data after training on GM12878, H1-ESC, and K562 (replicate1) separately. (b) APA plots of identified chromatin loops in GM12878, H1-ESC, and K562 (replicate1). (c) Venn diagram of DLoopCaller identified chromatin loops determined by CTCF ChIA-PET experiments in GM12878, H1-ESC and K562 (replicate1). (d) The proportion of CTCF ChIA-PET interactions types for H1-ESC and K562. The proportion of identified chromatin loops types using CTCF ChIA-PET data after training for H1-ESC and K562 (replicate1).

<https://doi.org/10.1371/journal.pcbi.1010572.g006>

correlations to estimate cooperativity and showed the strongest enrichment between KLF5 and ZN700, KLF3, KLF6, SP2, and SP1 motifs in H1-ESC (Fig 7). Moreover, the cooperativity estimations in GM12878 and K562 (replicate1) were shown in S8 and S9 Figs, it is obvious that the strongest enrichment between ZN770 and PAX5, and ZN121 motifs, between IKZF1 and PAX5 in K562, strongest enrichment between ZN700 and E2F7, and ZSC22 motifs in GM12878. This phenomenon suggests that motif enrichment differs in the identified

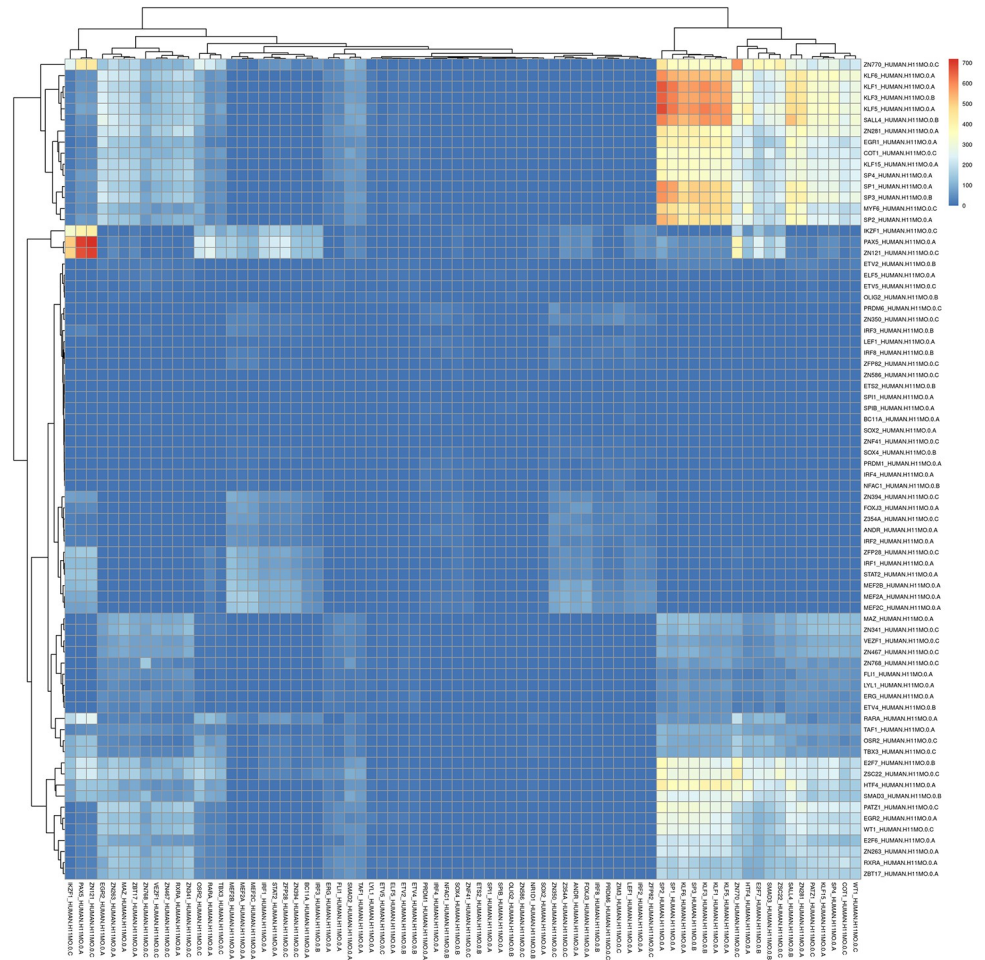


Fig 7. The co-enrichment of transcription factors on identified chromatin loops in H1-ESC with CTCF ChIA-PET training model.

<https://doi.org/10.1371/journal.pcbi.1010572.g007>

chromatin loops from different cell lines, which may provide helpful analysis of transcriptional regulation.

In addition, we find that the majority of the SMC1 HiChIP loops in mESC are distributed within 250kb accounting for up to 90.4% (13288/14695), and the APA plots show that the enrichment is more obvious in the lower left, which is similar to H1-ESC (S10(a)-(b) Fig). As shown in S6(C) Fig, and the overlap ratio of SMC1 HiChIP loops in mESC and GM12878 is 13.2% (2424/18332) even mismatches between either anchor are allowed, which suggests that the identified chromatin loops are specific across species even using the same enrichment technology. Moreover, from S11 and S12 Figs, the transcription factor motif co-enrichment in GM12878 and mESC demonstrates specificity. We conclude that the identified chromatin loops exhibit specificity for significant transcription factor motif co-enrichment across different cell lines and species.

The reproducibility and robustness of DLoopCaller

We also evaluated the degree of repeatability and robustness of DLoopCaller across biological replicates and different sequencing depth. We performed DLoopCaller on two replicates of K562 with CTCF ChIA-PET training to verify the reproducibility of DLoopCaller. Due to the

large difference in mapping read size between the two replicates, we used different thresholds to keep the similar number of chromatin loops identified in the two replicates. As shown in [S13 Fig](#), with regard to the distance distribution, APA analysis, and the proportion of regulatory elements of identified chromatin loops exhibit similarity in the two replicates. And it is obvious that the common strongest enrichment between ZN770 and PAX5, and ZN121 motifs in the two replicates ([S8](#) and [S14 Figs](#)). Overall, the chromatin loops identified by DLoopCaller in the two replicates are similar to some extent, which proves that DLoopCaller is reproducible. In addition, we adopted a binomial probability used in Peakachu to downsample the contact map of GM12878 without re-mapping, and we performed DLoopCaller with H3K27ac HiChIP training on the 80%, 50% and 30% down-sampled matrix respectively with 1.6 billion, 1 billion and 600 million cis-reads. The experimental results show that the identified loops at different sequencing depths maintain a large degree of overlap with those on original hic matrix, especially 77.1%, 72.2% and 76.4% in the 80%, 50% and 30% down-sampled matrix ([S15 Fig](#)), which indicates DLoopCaller is robust.

Discussion

With the rapid development of chromatin conformation capture technologies, which provides opportunities to dissect the role of the 3D structure of chromatin in cellular processes, including regulation of gene expression and DNA replication. Here, we proposed a novel method DLoopCaller integrating Hi-C contact maps and accessible chromatin landscape data to identify genome-wide chromatin loops. The main contribution of DLoopCaller lies in the following points: (i) We used the chromatin landscape data to generate a chromatin landscape matrix that matches the Hi-C contact maps, avoiding manual feature extraction; (ii) applied massive enriched experimental data, such as ChIA-PET/HiChIP and Capture Hi-C, to annotate positive samples; (iii) developed a deep learning framework to simultaneously extract features from Hi-C matrix and accessible chromatin landscape matrix to improve the accuracy of identifying chromatin loops in the whole genome. The experimental results show that DLoopCaller can effectively improve the accuracy of identifying chromatin loops compared with competing methods and identify a series of unique chromatin loops. We find that the identified chromatin loops from H3K27ac HiChIP contain more short-range loops while the ones from promoter Capture Hi-C contain more long-range loops in GM12878. Moreover, we discovery that most of the chromatin loops identified by DLoopCaller are mediated by enhancers, which is largely consistent with the used enrichment experimental data. Next, according to the analysis of the experimental results, identified chromatin loops show cell type specificity with low overlapping ratio across cell lines. Then significant transcription factor motif co-enrichment in identified chromatin loops exhibits specificity across different cell lines and across species. Last but not least, DLoopCaller is reproducible and robust across different biological replicates and sequencing depths.

Although DLoopCaller has achieves excellent performance and makes some new discoveries, there are still many limitations: (i) The training data of DLoopCaller depends on the enrichment experimental data, but the sources of different batches will produce noise in the experimental results; (ii) although DLoopCaller identifies many unique chromatin loops, they may have false positives; (ii) the adopted deep learning framework is a black-box model, which is difficult to interpret the extracted features for the identification of chromatin loops. Despite the above limitations, there are still a lot of works to be worth pursuing: (i) We perform DLoopCaller on more Hi-C data obtained from different chromatin conformation capture technologies to verify the effectiveness of the method, such as DNA SPRITE data and Micro-C maps; (ii) The results obtained by our method may be used to predict enhancer-promoter interactions to help further understanding of transcriptional regulatory mechanisms, which

remains a major challenge; (iii) to date, there are some pioneering works to enhance the resolution of Hi-C data, such as HiCPlus [50], HiCNN [51], and DeepHiC [52], we could try to perform DLoopCaller on high resolution Hi-C matrices to reduce false positives; (iv) we could try to incorporate more one-dimensional chromatin maps, such as histone modification data and gene expression data, to further improve the accuracy of identifying chromatin loops; (v) improving the efficiency of predicting chromatin loops on the whole genome.

Supporting information

S1 Note. The definition of evaluation metrics.

(DOCX)

S1 Table. The number of samples in each dataset.

(DOCX)

S2 Table. The detailed settings of DLoopCaller.

(DOCX)

S1 Fig. The line charts about F1-score and PRAUC for all chromosomes in GM12878, K562 (replicate1), H1-ESC and mESC.

(TIF)

S2 Fig. The boxplots about F1-score and PRAUC for all chromosomes in GM12878, K562 (replicate1), H1-ESC and mESC.

(TIF)

S3 Fig. The boxplot about Precision and Recall for all chromosomes in GM12878, K562 (replicate1), H1-ESC and mESC.

(TIF)

S4 Fig. APA plots for CTCF ChIA-PET, H3K27ac HiChIP, SMC1 HiChIP, RAD ChIA-PET, and promoter Capture Hi-C loops in GM12878 cell lines.

(TIF)

S5 Fig. (a) Venn diagram of CTCF ChIA-PET chromatin loops determined by DLoopCaller and Peakachu in GM12878; (b) Venn diagram of H3k27ac HiChIP chromatin loops determined by DLoopCaller and Peakachu in GM12878; (c) Distance distribution of Peakachu identified chromatin loops from Hi-C contact maps by using CTCF ChIA-PET data after training on GM12878; (d) Distance distribution of Peakachu identified chromatin loops from Hi-C contact maps by using H3k27ac HiChIP data after training on GM12878.

(TIF)

S6 Fig. Visual examples of identified loops by different models in a region. The black dots in the upper half of the three diamond-shaped graphs represent the chromatin loops identified by DLoopCaller, and the blue, green, and yellow dots in the lower half represent the chromatin loops identified by Peakachu, Fit-HiC, and HiCCUPS respectively.

(TIF)

S7 Fig. The cell type-specific loops with unique cell type specific chromatin accessibility or histone modification features.

(TIF)

S8 Fig. The co-enrichment of transcription factor on identified chromatin loops in K562 (replicate1) with CTCF ChIA-PET training model.

(TIF)

S9 Fig. The co-enrichment of transcription factor on identified chromatin loops in GM12878 with CTCF ChIA-PET training model.

(TIF)

S10 Fig. (a) Distance distribution of DLoopCaller identified SMC1 HiChIP chromatin loops from Hi-C contact maps in mESC. (b) The APA plots for SMC1 HiChIP chromatin loops in mESC. (c) Venn diagram of DLoopCaller identified SMC1 HiChIP chromatin loops in GM12878 and mESC.

(TIF)

S11 Fig. The co-enrichment of transcription factor on identified chromatin loops in GM12878 with SMC1 HiChIP training model.

(TIF)

S12 Fig. The co-enrichment of transcription factor on identified chromatin loops in mESC with SMC1 HiChIP training model.

(TIF)

S13 Fig. (a) Distance distribution of DLoopCaller identified chromatin loops from Hi-C contact maps by using CTCF ChIA-PET data after training on two replicates of K562; (b) APA plots for DLoopCaller CTCF ChIA-PET loops in two replicates of K562; (c) The proportion of identified chromatin loops types using CTCF ChIA-PET data after training for two replicates of K562; (d) Venn diagram of CTCF ChIA-PET chromatin loops in two replicates of K562.

(TIF)

S14 Fig. The co-enrichment of transcription factor on identified chromatin loops in K562 (replicate2) with CTCF ChIA-PET training model.

(TIF)

S15 Fig. Concordance of identified loops from datasets different down-sampled rates.

(TIF)

Author Contributions**Conceptualization:** Siguo Wang, Qinhu Zhang, De-Shuang Huang.**Data curation:** Ying He, Zhenghao Guo.**Formal analysis:** Siguo Wang.**Investigation:** Siguo Wang, Ying He, Zhenghao Guo.**Methodology:** Siguo Wang, Qinhu Zhang.**Validation:** Kyungsook Han.**Visualization:** Zhen Cui.**Writing – original draft:** Siguo Wang, Qinhu Zhang.**Writing – review & editing:** Siguo Wang, Qinhu Zhang, Ying He, De-Shuang Huang.**References**

1. Mumbach MR, Satpathy AT, Boyle EA, Dai C, Gowen BG, Cho SW, et al. Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. *Nature genetics*. 2017; 49(11):1602–12. <https://doi.org/10.1038/ng.3963> PMID: 28945252

2. Wang E, Zaman N, Mcgee S, Milanese J-S, Masoudi-Nejad A, O'Connor-McCourt M, editors. Predictive genomics: a cancer hallmark network framework for predicting tumor clinical phenotypes using genome sequencing data. *Seminars in cancer biology*; 2015: Elsevier.
3. Lee W, Huang D-S, Han K. Constructing cancer patient-specific and group-specific gene networks with multi-omics data. *BMC medical genomics*. 2020; 13(6):1–12. <https://doi.org/10.1186/s12920-020-00736-7> PMID: 32854705
4. Dekker J, Marti-Renom MA, Mirny LA. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nature Reviews Genetics*. 2013; 14(6):390–403. <https://doi.org/10.1038/nrg3454> PMID: 23657480
5. Dixon JR, Jung I, Selvaraj S, Shen Y, Antosiewicz-Bourget JE, Lee AY, et al. Chromatin architecture reorganization during stem cell differentiation. *Nature*. 2015; 518(7539):331–6. <https://doi.org/10.1038/nature14222> PMID: 25693564
6. Gorkin DU, Leung D, Ren B. The 3D genome in transcriptional regulation and pluripotency. *Cell stem cell*. 2014; 14(6):762–75. <https://doi.org/10.1016/j.stem.2014.05.017> PMID: 24905166
7. 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*. 2009; 326(5950):289–93. <https://doi.org/10.1126/science.1181369> PMID: 19815776
8. Rao SS, 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*. 2014; 159(7):1665–80. <https://doi.org/10.1016/j.cell.2014.11.021> PMID: 25497547
9. 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*. 2012; 485(7398):376–80. <https://doi.org/10.1038/nature11082> PMID: 22495300
10. Levine M. Transcriptional enhancers in animal development and evolution. *Current Biology*. 2010; 20(17):R754–R63. <https://doi.org/10.1016/j.cub.2010.06.070> PMID: 20833320
11. Ji X, Dadon DB, Powell BE, Fan ZP, Borges-Rivera D, Shachar S, et al. 3D chromosome regulatory landscape of human pluripotent cells. *Cell stem cell*. 2016; 18(2):262–75. <https://doi.org/10.1016/j.stem.2015.11.007> PMID: 26686465
12. Yuan L, Guo L-H, Yuan C-A, Zhang Y, Han K, Nandi AK, et al. Integration of multi-omics data for gene regulatory network inference and application to breast cancer. *IEEE/ACM transactions on computational biology and bioinformatics*. 2018; 16(3):782–91. <https://doi.org/10.1109/TCBB.2018.2866836> PMID: 30137012
13. Tang Z, Luo OJ, Li X, Zheng M, Zhu JJ, Szalaj P, et al. CTCF-mediated human 3D genome architecture reveals chromatin topology for transcription. *Cell*. 2015; 163(7):1611–27. <https://doi.org/10.1016/j.cell.2015.11.024> PMID: 26686651
14. Sanborn AL, Rao SS, Huang S-C, Durand NC, Huntley MH, Jewett AI, et al. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proceedings of the National Academy of Sciences*. 2015; 112(47):E6456–E65.
15. Rudan MV, Barrington C, Henderson S, Ernst C, Odom DT, Tanay A, et al. Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell reports*. 2015; 10(8):1297–309. <https://doi.org/10.1016/j.celrep.2015.02.004> PMID: 25732821
16. Quinodoz SA, Ollikainen N, Tabak B, Palla A, Schmidt JM, Detmar E, et al. Higher-order inter-chromosomal hubs shape 3D genome organization in the nucleus. *Cell*. 2018; 174(3):744–57. e24. <https://doi.org/10.1016/j.cell.2018.05.024> PMID: 29887377
17. Arrastia MV, Jachowicz JW, Ollikainen N, Curtis MS, Lai C, Quinodoz SA, et al. A single-cell method to map higher-order 3D genome organization in thousands of individual cells reveals structural heterogeneity in mouse ES cells. *bioRxiv*. 2020.
18. Jäger R, Migliorini G, Henrion M, Kandaswamy R, Speedy HE, Heindl A, et al. Capture Hi-C identifies the chromatin interactome of colorectal cancer risk loci. *Nature communications*. 2015; 6(1):1–9. <https://doi.org/10.1038/ncomms7178> PMID: 25695508
19. Fullwood MJ, Wei C-L, Liu ET, Ruan Y. Next-generation DNA sequencing of paired-end tags (PET) for transcriptome and genome analyses. *Genome research*. 2009; 19(4):521–32. <https://doi.org/10.1101/gr.074906.107> PMID: 19339662
20. 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*. 2016; 13(11):919–22. <https://doi.org/10.1038/nmeth.3999> PMID: 27643841
21. Rust MJ, Bates M, Zhuang X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature methods*. 2006; 3(10):793–6. <https://doi.org/10.1038/nmeth929> PMID: 16896339

22. Gustafsson MG. Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. *Journal of microscopy*. 2000; 198(2):82–7. <https://doi.org/10.1046/j.1365-2818.2000.00710.x> PMID: 10810003
23. Ay F, Bailey TL, Noble WS. Statistical confidence estimation for Hi-C data reveals regulatory chromatin contacts. *Genome research*. 2014; 24(6):999–1011. <https://doi.org/10.1101/gr.160374.113> PMID: 24501021
24. Durand NC, Shamim MS, Machol I, Rao SS, Huntley MH, Lander ES, et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell systems*. 2016; 3(1):95–8. <https://doi.org/10.1016/j.cels.2016.07.002> PMID: 27467249
25. Cairns J, Freire-Pritchett P, Wingett SW, Várnai C, Dimond A, Plagnol V, et al. CHiCAGO: robust detection of DNA looping interactions in Capture Hi-C data. *Genome biology*. 2016; 17(1):1–17. <https://doi.org/10.1186/s13059-016-0992-2> PMID: 27306882
26. Al Bkhetan Z, Plewczynski D. Three-dimensional epigenome statistical model: genome-wide chromatin looping prediction. *Scientific reports*. 2018; 8(1):1–11.
27. Kai Y, Andricovich J, Zeng Z, Zhu J, Tzatsos A, Peng W. Predicting CTCF-mediated chromatin interactions by integrating genomic and epigenomic features. *Nature communications*. 2018; 9(1):1–14.
28. Chen L, Capra JA. Learning and interpreting the gene regulatory grammar in a deep learning framework. *PLoS computational biology*. 2020; 16(11):e1008334. <https://doi.org/10.1371/journal.pcbi.1008334> PMID: 33137083
29. Leung MK, Xiong HY, Lee LJ, Frey BJ. Deep learning of the tissue-regulated splicing code. *Bioinformatics*. 2014; 30(12):i121–i9. <https://doi.org/10.1093/bioinformatics/btu277> PMID: 24931975
30. Chen Y, Li Y, Narayan R, Subramanian A, Xie X. Gene expression inference with deep learning. *Bioinformatics*. 2016; 32(12):1832–9. <https://doi.org/10.1093/bioinformatics/btw074> PMID: 26873929
31. Wang S, He Y, Chen Z, Zhang Q. FCNGRU: Locating Transcription Factor Binding Sites by combing Fully Convolutional Neural Network with Gated Recurrent Unit. *IEEE Journal of Biomedical and Health Informatics*. 2021.
32. Zhang Q, Wang S, Chen Z, He Y, Liu Q, Huang D-S. Locating transcription factor binding sites by fully convolutional neural network. *Briefings in bioinformatics*. 2021; 22(5):bbaa435. <https://doi.org/10.1093/bib/bbaa435> PMID: 33498086
33. Lv H, Dao F-Y, Zulfiqar H, Su W, Ding H, Liu L, et al. A sequence-based deep learning approach to predict CTCF-mediated chromatin loop. *Briefings in bioinformatics*. 2021; 22(5):bbab031. <https://doi.org/10.1093/bib/bbab031> PMID: 33634313
34. Trieu T, Martinez-Fundichely A, Khurana E. DeepMILO: a deep learning approach to predict the impact of non-coding sequence variants on 3D chromatin structure. *Genome biology*. 2020; 21(1):1–11. <https://doi.org/10.1186/s13059-020-01987-4> PMID: 32216817
35. Roayaei Ardakany A, Gezer HT, Lonardi S, Ay F. Mustache: multi-scale detection of chromatin loops from Hi-C and Micro-C maps using scale-space representation. *Genome biology*. 2020; 21(1):1–17. <https://doi.org/10.1186/s13059-020-02167-0> PMID: 32998764
36. Salameh TJ, Wang X, Song F, Zhang B, Wright SM, Khunsriaksakul C, et al. A supervised learning framework for chromatin loop detection in genome-wide contact maps. *Nature communications*. 2020; 11(1):1–12.
37. Heidari N, Phanstiel DH, He C, Grubert F, Jahanbani F, Kasowski M, et al. Genome-wide map of regulatory interactions in the human genome. *Genome research*. 2014; 24(12):1905–17. <https://doi.org/10.1101/gr.176586.114> PMID: 25228660
38. Alipanahi B, DeLong A, Weirauch MT, Frey BJ. Predicting the sequence specificities of DNA-and RNA-binding proteins by deep learning. *Nature biotechnology*. 2015; 33(8):831–8. <https://doi.org/10.1038/nbt.3300> PMID: 26213851
39. Zhang Q, Shen Z, Huang D-S. Predicting in-vitro transcription factor binding sites using DNA sequence + shape. *IEEE/ACM transactions on computational biology and bioinformatics*. 2019; 18(2):667–76.
40. He Y, Shen Z, Zhang Q, Wang S, Huang D-S. A survey on deep learning in DNA/RNA motif mining. *Briefings in Bioinformatics*. 2021; 22(4):bbaa229. <https://doi.org/10.1093/bib/bbaa229> PMID: 33005921
41. Zhang Q, He Y, Wang S, Chen Z, Guo Z, Cui Z, et al. Base-resolution prediction of transcription factor binding signals by a deep learning framework. *PLoS computational biology*. 2022; 18(3):e1009941. <https://doi.org/10.1371/journal.pcbi.1009941> PMID: 35263332
42. Glorot X, Bengio Y, editors. Understanding the difficulty of training deep feedforward neural networks. *Proceedings of the thirteenth international conference on artificial intelligence and statistics; 2010: JMLR Workshop and Conference Proceedings*.
43. Fullwood MJ, Liu MH, Pan YF, Liu J, Xu H, Mohamed YB, et al. An oestrogen-receptor- α -bound human chromatin interactome. *Nature*. 2009; 462(7269):58–64.

44. Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature genetics*. 2007; 39(3):311–8. <https://doi.org/10.1038/ng1966> PMID: 17277777
45. Tang L, Hill MC, Wang J, Wang J, Martin JF, Li M. Predicting unrecognized enhancer-mediated genome topology by an ensemble machine learning model. *Genome research*. 2020; 30(12):1835–45. <https://doi.org/10.1101/gr.264606.120> PMID: 33184104
46. Sanyal A, Lajoie BR, Jain G, Dekker J. The long-range interaction landscape of gene promoters. *Nature*. 2012; 489(7414):109–13. <https://doi.org/10.1038/nature11279> PMID: 22955621
47. Pliner HA, Packer JS, McFaline-Figueroa JL, Cusanovich DA, Daza RM, Aghamirzaie D, et al. Cicero predicts cis-regulatory DNA interactions from single-cell chromatin accessibility data. *Molecular cell*. 2018; 71(5):858–71. e8. <https://doi.org/10.1016/j.molcel.2018.06.044> PMID: 30078726
48. Yang Y, Zhang R, Singh S, Ma J. Exploiting sequence-based features for predicting enhancer–promoter interactions. *Bioinformatics*. 2017; 33(14):i252–i60. <https://doi.org/10.1093/bioinformatics/btx257> PMID: 28881991
49. Hammelman J, Krismer K, Gifford DK. spatzie: An R package for identifying significant transcription factor motif co-enrichment from enhancer-promoter interactions. *bioRxiv*. 2021.
50. Zhang Y, An L, Xu J, Zhang B, Zheng WJ, Hu M, et al. Enhancing Hi-C data resolution with deep convolutional neural network HiCPlus. *Nature communications*. 2018; 9(1):1–9.
51. Liu T, Wang Z. HiCNN: a very deep convolutional neural network to better enhance the resolution of Hi-C data. *Bioinformatics*. 2019; 35(21):4222–8. <https://doi.org/10.1093/bioinformatics/btz251> PMID: 31056636
52. Hong H, Jiang S, Li H, Du G, Sun Y, Tao H, et al. DeepHiC: A generative adversarial network for enhancing Hi-C data resolution. *PLoS computational biology*. 2020; 16(2):e1007287. <https://doi.org/10.1371/journal.pcbi.1007287> PMID: 32084131