

## REVIEW

# Computational methods for identifying enhancer-promoter interactions

Haiyan Gong<sup>1</sup>, Zhengyuan Chen<sup>1</sup>, Yuxin Tang<sup>1</sup>, Minghong Li<sup>1</sup>, Sichen Zhang<sup>1</sup>, Xiaotong Zhang<sup>1,3,\*</sup>, Yang Chen<sup>2,\*</sup>

<sup>1</sup> School of Computer and Communication Engineering, Beijing Advanced Innovation Center for Materials Genome Engineering, University of Science and Technology Beijing, Beijing100083, China

<sup>2</sup> State Key Laboratory of Medical Molecular Biology, Department of Biochemistry and Molecular Biology, Institute of Basic Medical Sciences, School of Basic Medicine, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing 100005, China

<sup>3</sup> Shunde Innovation School, University of Science and Technology Beijing, Foshan 528399, China

\* Correspondence: [yc@ibms.pumc.edu.cn](mailto:yc@ibms.pumc.edu.cn); [zxt@ies.ustb.edu.cn](mailto:zxt@ies.ustb.edu.cn)

Received September 14, 2022; Revised October 10, 2022; Accepted October 17, 2022

**Background:** As parts of the cis-regulatory mechanism of the human genome, interactions between distal enhancers and proximal promoters play a crucial role. Enhancers, promoters, and enhancer-promoter interactions (EPIs) can be detected using many sequencing technologies and computation models. However, a systematic review that summarizes these EPI identification methods and that can help researchers apply and optimize them is still needed.

**Results:** In this review, we first emphasize the role of EPIs in regulating gene expression and describe a generic framework for predicting enhancer-promoter interaction. Next, we review prediction methods for enhancers, promoters, loops, and enhancer-promoter interactions using different data features that have emerged since 2010, and we summarize the websites available for obtaining enhancers, promoters, and enhancer-promoter interaction datasets. Finally, we review the application of the methods for identifying EPIs in diseases such as cancer.

**Conclusions:** The advance of computer technology has allowed traditional machine learning, and deep learning methods to be used to predict enhancer, promoter, and EPIs from genetic, genomic, and epigenomic features. In the past decade, models based on deep learning, especially transfer learning, have been proposed for directly predicting enhancer-promoter interactions from DNA sequences, and these models can reduce the parameter training time required of bioinformatics researchers. We believe this review can provide detailed research frameworks for researchers who are beginning to study enhancers, promoters, and their interactions.

**Keywords:** enhancer; promoter; enhancer-promoter interaction; machine learning; deep learning

**Author summary:** In this review article, we discuss prediction methods for enhancers, promoters, and enhancer-promoter interactions that may help researchers choose methods for identifying cis-regulatory elements.

## INTRODUCTION

It is known that cis-acting regulatory elements (CREs) are DNA sequences that have transcriptional regulatory functions in the human genome. An enhancer (20- to 400-bp) [1] is a class of non-coding DNA sequences bound by transcription factors [2], and these sequences can interact with short regions of DNA (100–1000 bp),

known as promoters, located near the gene transcription start sites (TSS) of a gene [3]. Enhancers and promoters are essential cis-regulatory elements for promoting gene transcription activities over a long distance. The interactions between distal enhancers (even with tens of kilobases) and proximal promoters regulate target genes and inhibit the cis-regulatory mechanism of the human genome [4–9].

Studying the mechanism of enhancer and promoter interactions (EPIs) may help us to understand the regulatory relationships among genes and reveal the genes associated with diseases. Davison *et al.* showed that EPIs can lead to type I diabetes and multiple sclerosis, and that new genes related to these diseases can be predicted using EPIs [10]. Smemo *et al.* [11] found the first intron region of the *FTO* gene in mice and humans, and the homologous gene *IRx3* was found to exist in a distal EPI. In the human brain, heart, and lungs high levels of *IRx3* gene are expressed; this is very important for controlling weight. Therefore, the study of EPIs, especially cell line-specific EPIs, may provide insight into the mechanisms of gene expression regulation, cell differentiation, and disease. In addition, research on EPIs has provided new methods and ideas for diagnosing and treating disease as well as for developing drugs.

Many sequencing technologies have been developed to generate data and identify enhancer, promoter, and chromosome interactions. For example, epigenomic features such as the histones and transcription factor binding sites (TFBS) data generated by chromatin immunoprecipitation (ChIP-seq) [12,13] and cleavage under targets and release using nuclease (CUT&RUN) [14] technologies have been widely used to identify enhancers and promoters. High-throughput chromosome conformation capture (Hi-C) [15] data (such as BL-Hi-C [16]) is frequently used to call loops (chromosome interactions that connect two distal regulatory elements). Promoter Capture Hi-C [17], Chromatin Interaction Analysis with Paired-End-Tag sequencing (ChIA-PET) [18], and HiChIP [19] can also identify genomic features such as enhancer-promoter interactions. Genetic features such as DNA sequences, pseudo dinucleotide composition (PseDNC), and Pseudo k-tuple nucleotide composition (PseKNC) [20] are also widely used to predict enhancers and promoters. Although the amount of high-throughput sequencing data is increasing rapidly, there are few enhancer-promoter interaction datasets that have been validated by experiments. The prediction of enhancer-promoter interactions using machine learning, deep learning, or other methods is therefore one of the most promising research topics in bioinformatics.

Numerous review articles have been published in recent decades concerning: enhancer interactions, including their role [21] at the genome-wide level; transcription enhancers in animal development, evolution [22], and disease [23]; functional contributions to transcription [24,25]; the functional significance of enhancer chromatin modification [26]; models that describe dynamic three-dimensional chromosome topology related to development enhancers; methods for

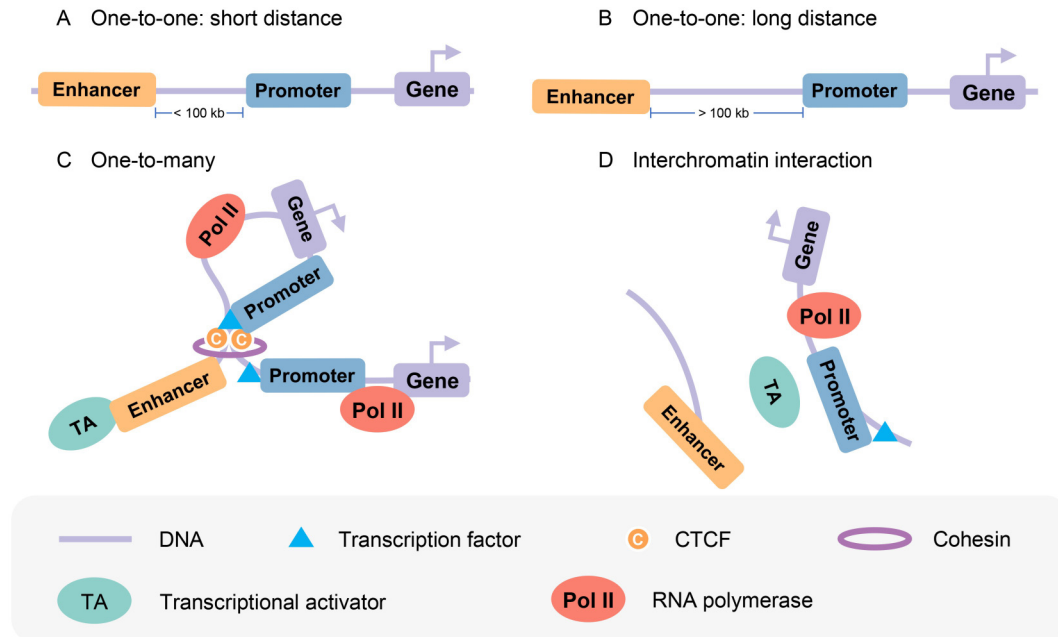
identifying enhancer target genes [27] and enhancers [28–30]; the mechanisms of EPIs in higher eukaryotes [31]; bioinformatics analysis methods related to EPIs prediction [32–35]; analysis from sequence data [36,37]; and how EPIs control gene expression [38]. However, with the advancement of computational methods in the past decade, research has increasingly proposed methods for detecting enhancer-promoter interaction tools based on traditional machine learning or deep learning, but there has yet to be a global overview of solutions specifically for EPI identification.

In light of this issue, this paper proposes computational models for identifying enhancer-promoter interactions based on high-throughput experimental data published from 2010 to 2022. First, we discuss the relationship between EPIs and gene transcription, and we provide a general framework for enhancer-promoter identification. Next, we discuss in detail recognition methods that have been developed in the last decade for enhancers and promoters, chromatin loops, and enhancer-promoter interactions; we summarize available enhancer and promoter resources, and suggest realistic guidelines for their use. Finally, we review the application of methods for identifying EPIs in diseases such as cancer.

## REGULATION OF GENE EXPRESSION VIA EPIs

Previous studies [39–41] have shown that the intrachromosomal and interchromosomal communications between enhancer and promoter regulate gene transcription. Transcription from target promoters can be activated by enhancers in interchromatin or intrachromatin over a short distance or a long distance (more than 100 kb) [1] (Fig. 1A, B), and one enhancer may interact with multiple promoters (Fig. 1C). He *et al.* [42] observed that the number of targets for each promoter is 2.92 on average. Some transcription factors may also mediate the interchromosomal interaction between enhancer and promoter (Fig. 1D). For example, Patel *et al.* [43] found a T-cell-specific cis-regulatory element in chromosome 16 (TIL16) that can interact with the *TAL1* promoter through interchromosomal interaction, and *c-Maf* and *p300* may cooperate to mediate the interchromosomal loop for abnormal activation of *TAL1* in T-ALL cells. Therefore, the prediction of enhancers, promoters, and their interactions is vital to our understanding of gene transcription mechanisms.

EPIs can be identified by formulating the problem as follows: “Given two DNA sequences (A and B) described by different data types, first, determine if either A or B can function as an enhancer or a promoter, then determine if A and B are a chromatin loop”. A



**Figure 1. Mechanisms of transcriptional activation over EPIs.** (A, B) Enhancers activate transcription from target promoters over a short distance or long distance. (C) One enhancer interacts with many promoters. (D) Interchromosomal enhancer and promoter interaction.

general process for identifying EPIs is shown in Fig. 2, which shows that the identification of EPIs can be divided into four categories:

(i) Given two DNA sequences with transcription factors (TFs), histone marks (provided by ChIP-seq), and chromatin interactions (provided by Hi-C) information, we first need to determine whether the given two DNA sequences are enhancers or promoters by calling peaks, or methods based on traditional machine learning, deep-learning. Then, we need to call loops from Hi-C data to determine whether the two DNA sequences form a chromatin loop.

(ii) Given two DNA sequences with protein chromatin interaction information (provided by ChIA-PET, HiChIP, or PChI-C), we can call chromatin loops to determine whether the two DNA sequences have EPIs.

(iii) Given two DNA sequences with TFs, histone marks features, and other epigenomic features, two DNA sequences can be identified as EPIs or not by machine-learning-based methods.

(iv) Given two DNA sequences without other information, the two DNA sequences can be identified as EPIs or not by deep-learning-based methods.

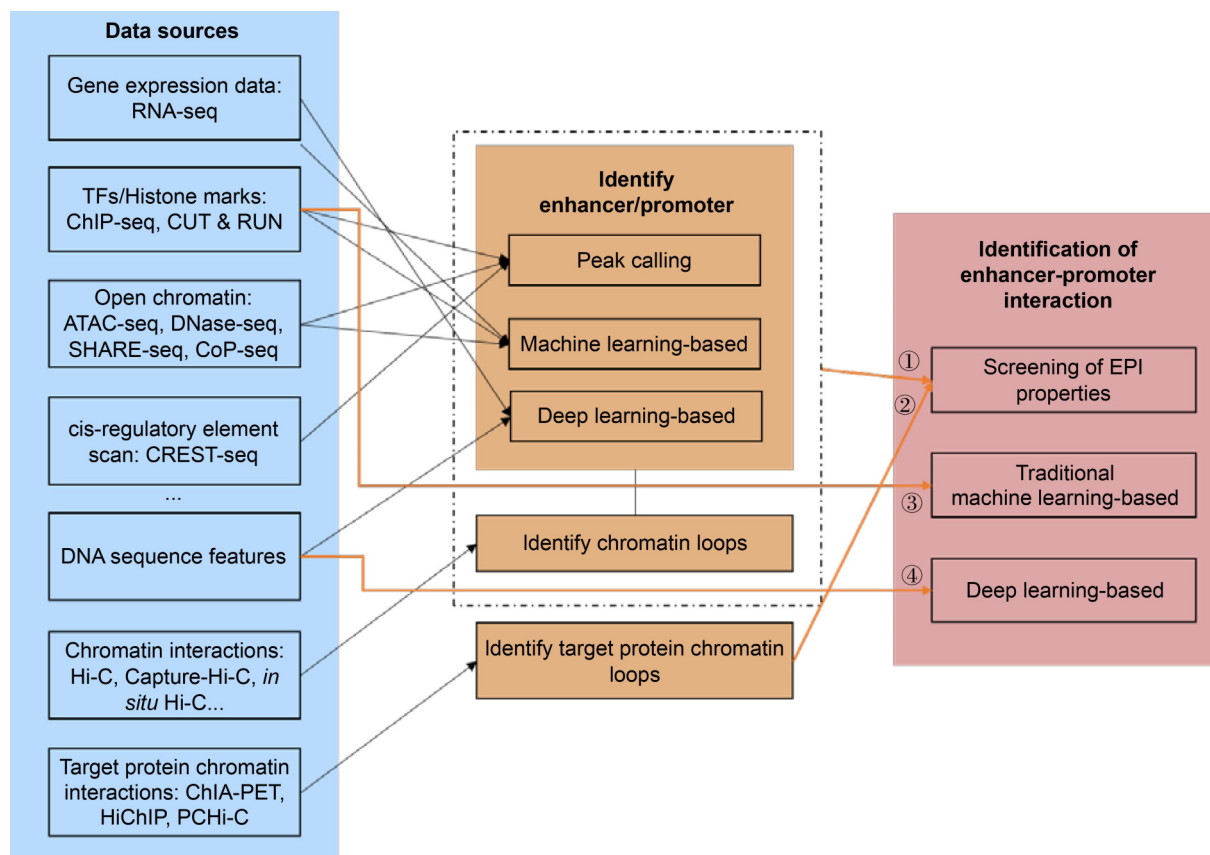
Thus we see that the data analysis process can be categorized into the prediction of enhancers, promoters, and EPIs. In the following sections, we describe the prediction of enhancers and promoters, and the identification of EPIs, separately.

## PREDICTION OF ENHANCER AND PROMOTER

As Table 1, Table 2, and Fig. 3 show, we can choose methods based on traditional machine learning or deep learning to check if a given DNA sequence is an enhancer or a promoter. To do this, we first need to process the DNA sequence, generate a training set with labels (promoter, enhancer, or none), and then identify enhancers or promoters by traditional machine learning or deep learning.

### Vector representations of DNA sequences

To generate the DNA sequence vectors that can be recognized by traditional machine learning or deep learning, we first need to code the DNA sequence (*e.g.*, ATCGGC...) in one of the following ways. (i) One-hot encoding, although has two problems: (1) the curse of dimensionality and (2) the distance between any pair of one-hot vectors is equal. (ii) To overcome the two problems of one-hot encoding, we use a word embedding algorithm, such as Word2vec [114] or Glove [115], to encode the DNA sequence. For example, dna2vec [116] first transforms a sequence into k-mers (a DNA sequence of length k) and then transforms the k-mers into vectors using Word2vec.



**Figure 2. An overview of the EPI prediction methods using different data sources.**

**Table 1 Prediction methods of enhancer**

Category	Refs.	Time	Source data	Method	Software name	Citation number
Traditional machine learning-based	[44]	2011	DNA sequence	SVM (support vector machine) classifier	k-mer-svm	162
	[45]	2012	TF motifs	LASSO regression	CLARE	9
	[46]	2012	ChIP-seq histone methylation and acetylation maps	Genetic algorithm-optimized support vector machine	ChromaGenSVM	77
	[47]	2013	Histone modification ChIP-seq	Random-forest- based	RFECS	144
	[48]	2014	Gapped k-mer features	SVM	gkm-svm	239
	[49]	2014	Histone modifications (ChIP-Seq), TFBSs, chromatin accessibility (DNase-Seq), transcription (RNA-Seq), evolutionary conservation, sequence signatures	Linear SVM and multiple kernel learning	EnhancerFinder	162
	[50]	2015	ChIP-seq	AdaBoost-based	DELTA	31
	[51]	2015	Histone ChIP-seq and DNA sequence	SVM	DEEP	73
	[52]	2016	DNA sequence	Machine learning	iEnhancer-PsedeKNC	15
	[53]	2016	DNA sequence, pseudo k-tuple nucleotide composition	SVM	iEnhancer-2L	334
	[54]	2016	Chromatin state, DNA sequence	A two-step wrapper- based feature selection method	EnhancerPred	52
	[55]	2016	WGBS DNA methylation profiles	Weighted support vector machine learning framework	LMethyR-SVM	9

*(continued)*

Category	Refs.	Time	Source data	Method	Software name	Citation number	
Traditional machine learning-based	[56]	2017	Short dinucleotide repeat motifs (DRMs), DNA sequence, enhancer-associated histone modification data	Machine learning	–	22	
	[57]	2017	Chromatin state, DNA sequence	A two-step wrapper-based feature selection method	EnhancerPred2.0	29	
	[58]	2017	Histone ChIP-seq and methylation, DNA sequence	Random forest	REPTILE	43	
	[59]	2018	DNA sequence	SVM	iEnhancer-EL	106	
	[60]	2018	FANTOM5 atlas of TrEns	Feature matrix generation, feature ranking using Gini-index, logistic regression	TELS	2	
	[61]	2018	DNA sequence	k-mer and machine learning based method	enhancer_prediction	13	
	[62]	2020	STARR-seq	Supervised machine-learning	MatchedFilter	21	
	[20]	2021	DNA sequence	Feature extraction technique and SVM	piEnPred	6	
	[63]	2021	Chromatin state and DNA sequence	Enhanced feature representation using random forest	iEnhancer-RF	8	
	[64]	2021	Nucleotide Composition	Two-Layer Predictor, Kullback-Leibler divergence, LASSO, SVM	iEnhancer-KL	1	
	[65]	2021	DNA sequence	7-mer and random forest	Computational CRISPR Strategy (CCS)	38	
	[66]	2021	DNA sequence	Random forest, extremely randomized tree, multilayer perceptron, SVM and extreme gradient boosting	Enhancer-IF	12	
	Deep learning-based	[67]	2010	Histone modification ChIP-seq	Time delay neural network (TDNN)	CSI-ANN	160
		[68]	2016	ChIP-Seq, DNase-Seq, RNA-Seq, DNA methylation, and other features	Deep learning-based	PEDLA	91
[69]		2017	DNA sequence	CNN (convolution neural network)	DeepEnhancer	76	
[70]		2017	DNA sequence	Deep-learning-based	BiRen	91	
[71]		2018	ATAC-Seq	Neural network-based model	PEAS	22	
[72]		2019	DNA sequence	Word embeddings and SVM	iEnhancer-5Step	96	
[73]		2020	DNA sequence	Word Embedding and CNN	iEnhancer-CNN	26	
[74]		2021	DNA sequence and DNase-seq	Deep-learning-based	DeepCAPE	8	
[75]		2021	STARR-seq	Deep-learning-based	DECODE	2	
[76]		2021	DNA sequence	Augmented data and Residual CNN	ES-ARCNN	4	
[77]		2021	Pseudo - K-tuple nucleotide composition and DNA sequence	DNN	iEnhancer-DHF	8	
[78]		2021	DNA sequence	Word embedding, generative adversarial net, CNN	iEnhancer-GAN	8	
[79]		2022	DNA sequence	Neural network	RicENN	1	
[80]		2022	DNA sequence	Enhanced feature extraction strategy, deep learning	–	0	
[81]		2022	DNA sequence	One-hot encoding, convolutional neural network	iEnhancer-Deep	2	
[82]		2022	DNA sequence	DBSCAN, random forest, word2vec and attention-based Bi-LSTM	–	0	

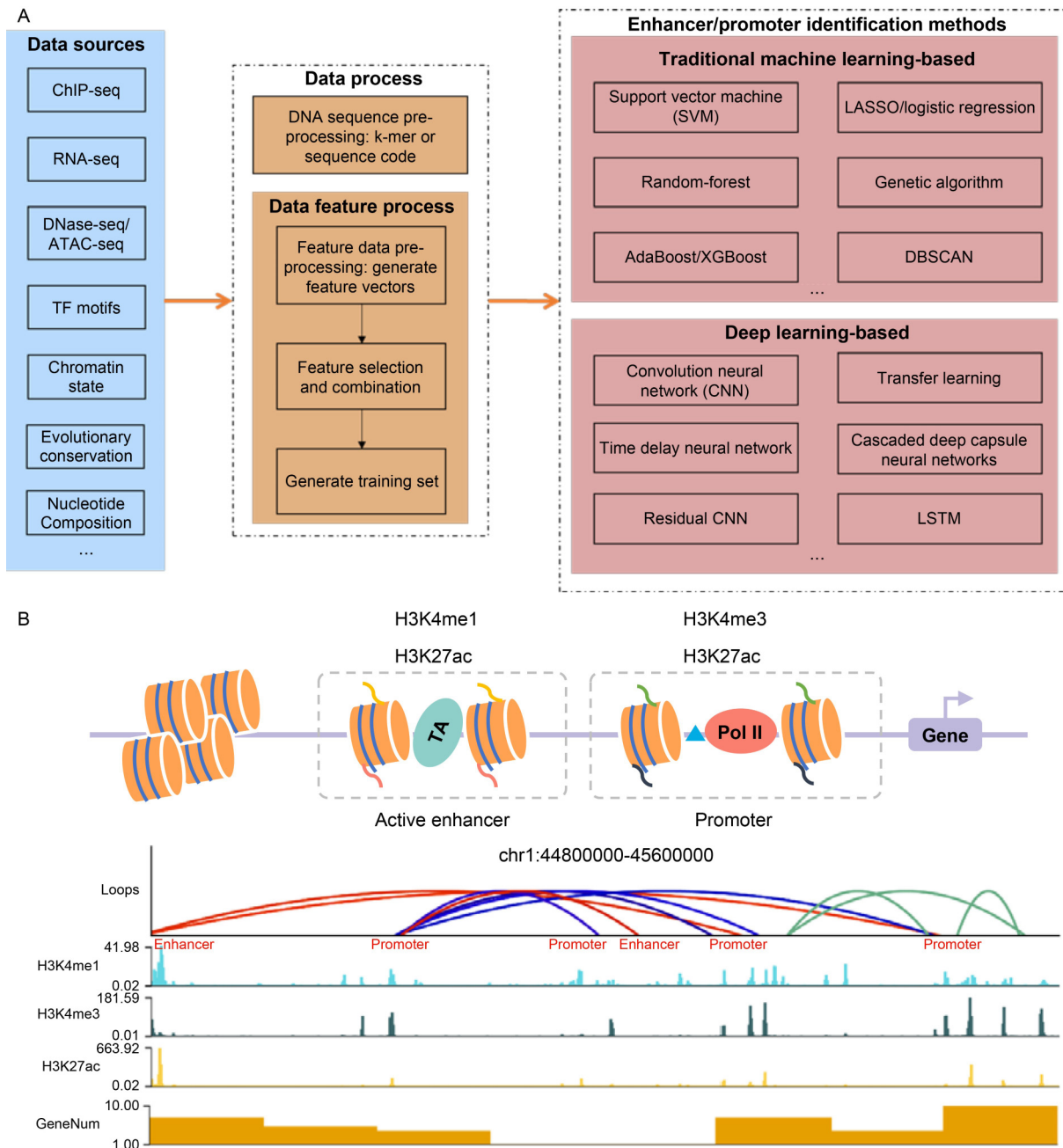
**Table 2 Prediction methods of promoter**

Category	Refs.	Time	Source data	Method	Software name	Citation number
Deep learning-based	[83]	2012	DNA sequence	DNA sequence features	–	63
	[84]	2016	DNA sequence	Deep feature selection, DFS		200
	[85]	2017	DNA sequence	CNN	CNNProm	169
	[86]	2018	DNA sequence	SVM	BacSVM+	9
	[87]	2018	DNA sequence	DNA sequence features	iPromoter-2L	256
	[88]	2019	DNA sequence	CNN and LSTM	DeePromoter	80
	[89]	2019	DNA sequence	Deep learning and combination of continuous FastText N-Grams	deepPromoter	46
	[90]	2019	DNA sequence	Deep learning	PromID	68
	[91]	2019	DNA sequence	Minimum redundancy maximum relevance (mRMR) algorithm and increment feature selection strategy, SVM	iProEP	99
	[92]	2019	DNA sequence	Combinee smoothing cutting window algorithm, k-mer, SVM	iPromoter-2L2.0	57
	[93]	2019	Bacterial $\sigma 70$ promoter sequences	Feature subspace based ensemble classifier	iPromoter-FSEn	30
	[94]	2019	Bacterial $\sigma 70$ promoter sequences	Multiple windowing and minimal features	iPro70-FMWin	20
	[95]	2019	The physicochemical properties of nucleotides and their nucleotide density into pseudo K-tuple nucleotide composition	A two-layer predictor	iPSW(2L)-PseKNC	55
	[96]	2019	DNA sequence	F-score feature selection method	MULTiPLY	87
	[97]	2020	DNA sequence of <i>Escherichia coli</i> K-12	Statistical physics model	PhysMPrePro	1
	[98]	2020	DNA sequence of <i>Escherichia coli</i> K-12	CNN	iPromoter-BnCNN	23
	[99]	2020	DNA sequence of <i>Escherichia coli</i> K-12	CNN, pseudo-di-nucleotide composition	PseDNC-DL	32
	[100]	2020	DNA sequence of <i>Escherichia coli</i> K-12	One-hot encoding and CNN	pePromoter-CNN	17
	[101]	2021	The k-mer nucleotide composition, binary encoding and dinucleotide property matrix-based distance	Extremely randomized trees	iPromoter-ET	5
	[102]	2021	Rice-specific DNA sequence	CNN	Cr-Prom	9
[103]	2021	DNA sequence of <i>Escherichia coli</i> K-12	A two-layer predictor	iPro2L-PSTKNC	5	
[104]	2021	DNA sequence	CNN	iPTT(2 L)-CNN	2	
[105]	2021	DNA sequence	Cascaded deep capsule neural networks	Depicter	23	
[106]	2022	DNA sequence	k-mers and deep learning network	PPred-PCKSM	1	
[107]	2022	DNA sequence	k-mer word vector, multiple descriptors and feature selection using XGBoost	dPromoter-XGBoost	1	
[108]	2022	DNA sequence	k-mers and LSTM network	–	1	
[109]	2022	DNA sequence	Moran-based spatial auto-cross correlation method and deep convolution generative adversarial network	iPro-GAN	2	
[110]	2022	Promoter data sets from both plants and humans	Synthetic sampling, transfer learning and label smoothing regularization	HMPI	0	
[111]	2022	Promoter sequences from six nanochloropsis strains	Densely connected convolutional neural networks	DenseNet-Predict Promoter	0	
Peak calling	[112]	2015	Capture Hi-C	–	–	861
	[113]	2016	Promoter capture Hi-C	–	–	769

The training sets for enhancers and promoters

There are two ways to obtain enhancer and promoter training sets: (i) Download data sets from a public data

repository. For example, we can download the human and mouse enhancer data sets from the SEDB [117] database and can download the eukaryotic promoter from the EPD [118] database. More databases for



**Figure 3. An overview of the enhancer and promoter identification process.** (A) EPI identification methods include data processing and identification methods. (B) An example of identifying EPIs based on Hi-C and ChIP-Seq data. The red lines represent EPIs, the blue lines represent promoter-promoter interactions, and the green lines represent other chromosome interactions.

enhancers and promoters are listed in Table 3. (ii) Available research has shown that H3K4me1 and H3K27ac enrichment occurs in both enhancers and promoters and that H3K4me1 together with H3K27ac, and a lack of H3K4me3 at the same genomic site can distinguish enhancers from promoters [49]. Additionally, enhancers are enriched with TFBS, Med1. Therefore, we can identify enhancers and promoters by

calling peaks from TFBS, H3K27ac, H3K4me1, H3K4me3, or Med1 ChIP-seq data. As Fig. 3B shows, we downloaded H3K27ac, H3K4me3, and H3K4me1 ChIP-seq data in the HeLa-S3 cell line from the ENCODE platform under accession number ENCSR-000AOC, ENCSR000AOF, and ENCSR000APW, respectively. The genome sites with H3K27ac, H3K4me3, and H3K4me1 ChIP-seq signals were

**Table 3 Available public data repository for enhancer and promoter**

Database type	Data repository name	
Enhancer	Sedb [117]	
	PRMod [119]	
	Human Transcribed Enhancer Atlas [120]	
	VISTA [121]	
	dbSUPER [122]	
	ENdb [123] (human enhancer)	
	SEA [124]	
	RAEdb [125]	
	SELER (human cancers) [126]	
	EnDisease [127]	
	dbInDel [128]	
	CancerEnD (cancer associated enhancers) [129]	
	CPE-DB [130]	
	Animal-eRNAdb [131]	
	Promoter	EPD [118]
		PlantProm (plant promoter) [132]
TransGene Promoters, TGP [133]		
Osteo-Promoter Database (OPD) skeletal cells [134]		
Osiris [135]		
TiProD [136]		
PromoterCAD (mammalian promoter/enhancer) [137]		
EPDNew [138]		
PPD [139]		

identified as promoters. The genome sites with H3K27ac, H3K4me1 CHIP-seq signals, but without H3K4me3 signals were identified as enhancers.

#### Methods for identifying enhancer/promoter based on traditional machine learning

In machine learning-based methods, the enhancer/promoter identification problem can be reformulated into a binary classification problem (yes or no). Since 2010, support vector machine (SVM) [20,44,46,48,49,51,53,55,59,64,66,86,91,92,140,141], regression [45,60], random forest [47,58,63,65,66,101], boost-based [50,66], and other traditional machine learning methods [52,56,61,62,83,84,87,93–96,103] have all been applied to predict enhancers and promoters. The SVM-based method combined with feature selection has been the most used, even within the last three years. For example, the kmer-SVM [44] first finds the motif related to enhancers by k-mer analysis, then inputs the motif into the SVM model to get the classification results. piEnPred [20] takes advantage of feature extraction

techniques such as k-mer, composition of k-spaced nucleic acid pairs (CKSNAP), Dinucleotide-based cross covariance (DCC), PseDNC, and PseKNC to extract features and SVM to classify enhancers and promoters.

Generally, there are three steps to traditional machine learning-based methods. (i) Use of feature extraction techniques to extract features [20,54,57,60,63,84,91,93,96], such as gene expression, histone modification marks, DNA sequence features, and TFs motifs. (ii) Classification of enhancers and promoters by classification algorithms, such as SVM, random forest, or regression. (iii) Tuning of the model parameters and optimization of the target functions using optimization algorithms, such as genetic algorithms [46].

After surveying the accession and citation numbers of these traditional machine-learning methods (Table 1), we recommend that users who do not want to run code using the web server iEnhancer-2L [53] should identify enhancers and their strengths using pseudo k-tuple nucleotide composition. For users who want to run code by themselves, we recommend that they choose gkm-svm [48], REPTILE [58], and CCS [65]. These tools provide detailed information and example data for users to get up to speed and run them quickly.

#### Methods for identifying enhancer/promoter based on deep-learning

Methods based on deep-learning primarily focus on training a neural network with DNA sequences or DNA sequences with epigenomic characteristics (such as histone modifications, chromatin accessibility, DNA methylation, or CpG islands) as inputs. Though some scholars have trained their networks with epigenome features [67,68,71,74,75,82], most have done so with only DNA sequences as inputs [69,70,72,73,77–81,85,88–90,98–100,102,104–111,142]. Predicting enhancers and promoters directly from DNA sequences is believed to be more applicable than identifying them from multiple epigenomic features because the epigenomic characteristics data carries with it substantial sequencing costs, and a high rate of false positives. However, prediction methods that use epigenomic characteristics in their inputs are more accurate than those that only use DNA sequences.

Methods based on deep-learning can be roughly divided into the following two steps. (i) Encoding a DNA sequence as in Section “Vector representations of DNA sequence”. (ii) Constructing a neural network to predict the presence of enhancers or promoters, such as CNN [69,73,76,78,81,85,88,98–100,102,104,111], transfer learning [110], or LSTM [82,88,108]. To establish the right characteristics and increase the accuracy of

identifying an enhancer or promoter, the above methods either improve the input layer of DNA feature vector representation (for example, dna2vec) or neural network architectures or change the activation functions. [Tables 1](#) and [2](#) list the available deep-learning-based methods for detecting enhancers and promoters. CSI-ANN [67] was the first deep learning-based method for the identification of enhancers, though Yang *et al.* [78] have since proposed iEnhancer-GAN to identify enhancers using word embedding, generative adversarial net, and CNN to capture DNA sequence features.

Although computational methods such as traditional machine learning and deep learning have achieved solid results, some problems still exist. One problem is that such methods typically use gene expression data such as chromatin characteristics and histone modification information as features to train models. When gene expression data are missing, these models cannot predict enhancers. Another problem is that enhancers are species-specific. That is, enhancers are expressed differently by different species, so the current methods have low performance in predicting enhancers across species.

For these deep-learning-based methods, we give some suggestions for tool selection. For users who want to predict using ChIP-seq, RNA-seq data, and other features as inputs, we recommend methods based on the input data requirements. For users who wish to identify enhancers and promoters with only DNA sequences as inputs, the number of citations metric ([Tables 1](#) and [2](#)) shows that BiRen [70] and PromID [90] are used frequently for predicting enhancers and promoters, respectively. Online tools including ES-ARCNN [76], iEnhancer-Deep [81] and iPromoter-2L [87] are easy to use and return the prediction results from these methods quickly.

## PREDICTION OF ENHANCER-PROMOTER INTERACTION

The task of recognizing EPIs is based on the prediction of enhancers and promoters individually in order to determine if there is an interaction between them, and this is a challenging task. First, multiple promoters can be activated by one enhancer, and multiple enhancers can coordinate to regulate one promoter. Secondly, EPI has tissue-specificity [42]. These features result in poor generalization for current EPI recognition methods. The existing EPI recognition methods are divided into three main types: (i) screen EPIs based on high-throughput sequencing experiments, (ii) methods based on traditional machine-learning, and (iii) methods based on deep-learning.

## Generation of EPIs training sets

In surveying the benchmarking EPI data sets used in 12 EPI identification methods ([Table 4](#)), we found 10 methods used the EPI data sets in GM12878, HUVEC, HeLa-S3, IMR90, K562, and NHEK cell line proposed by TargetFinder [143]. TargetFinder integrates TFs, histone markers, Dnase-seq, gene expression, and DNA methylation data to predict EPIs. However, before training any model, the EPI data sets need to be augmented, such as with the synthetic minority oversampling technique [156], because of the low ratio of positive to negative data sets (1/35). There are two ways to generate an acceptable EPI dataset.

(i) We can label the active enhancer and promoter regions using ChIP-seq data or annotation files and then annotate chromosome interactions from Hi-C data. For example, EPIP [154] obtained the enhancer data sets and identified the promoter data sets from transcription start site (TSS) annotation files by considering the genomic regions between the 1000 bases upstream and 100 bases downstream of the TSS regions. We can also obtain enhancer and promoter data sets from databases listed in [Table 3](#). To train an EPI identification model, we can divide the training dataset into positive and negative EPI data sets by overlapping the training data set with the regions of the loops called from Hi-C data [15]. For example, EPIP [154] states that if an enhancer and a promoter overlap with a pair of regions from loops within 30 reads, this pair of enhancer and promoter is considered a positive EPI. We can then use the loop callers listed in [Table 5](#) to call loops from Hi-C data, such as HiCCUPS [157], HiGlass [159], cLoops [160], FitHiC2 [161], Mustache [162], and HiC-ACT [164]. As [Fig. 3B](#) displays, to show how to identify EPIs, we downloaded the Hi-C data from 4dnucleome platform under accession number 4DNESCMX7L58, called loops using Mustache [162], and then annotated these loops as enhancer-promoter interactions or promoter-promoter interactions based on ChIP-seq signals.

(ii) We can also obtain EPI data sets by screening loops from target proteins HiChIP, PLAC-seq, or ChIA-PET data. For example, first, H3K27ac HiChIP data can be used to identify enhancer regions by calling loops. Then, we can screen loops that interact with promoters as EPIs. Many available loop callers have been developed for HiChIP, PLAC-seq, and ChIA-PET data. As [Table 5](#) shows, tools such as HiC-Pro [158], hichipper [166], MAPS [169], FitHiChIP [167], and HiChIP-Peaks [170] have been developed for HiChIP and PLAC-seq data, and tools like ChIA-PET Tool [171], MICC [173], ChIA-PET2 [175], ChIAPoP [176],

**Table 4 The benchmarking enhancer-promoter interaction dataset used in EPIs identification methods**

EPI dataset	EPIs methods that used the dataset
EPI Dataset provided by Whalen <i>et al.</i> [143]	PEP [144], EP2vec [145], SPEID [146], random forest based method [147], Zhuang <i>et al.</i> [148], EPIVAN [149], Singh <i>et al.</i> [150], EPI-DLMH [151], EPIsHilbert [152], EPI-Mind [153]
Dataset provided by Talukder <i>et al.</i> [154]	EPIP [154]
Dataset provided by Jing <i>et al.</i> [155]	SEPT [155]

**Table 5 Methods of calling loops from 3C-based data**

Publication	Time	Sequencing technology	Method	Software name	Citation number
[157]	2014	Hi-C [15]	Identify “enriched pixels” where the interaction frequency is higher than expected	HiCCUPS	753
[158]	2015	Hi-C, HiChIP	Toolkit	HiC-Pro	1125
[159]	2018	Hi-C		HiGlass	402
[160]	2020	Hi-C, ChIA-PET	DBSCAN-based	cLoops	35
[161]	2020	Hi-C	Identify loops from high-resolution Hi-C	FitHiC2	72
[162]	2020	Hi-C, Micro-C [163]	Scale-space representation	Mustache	42
[164]	2021	Hi-C	Aggregated Cauchy test	HiC-ACT	10
[165]	2021	Hi-C	Identify loops from high-resolution Hi-C	HiCORE	1
[166]	2018	HiChIP [19]	DNA loop calling	hichipper	86
[167]	2019	HiChIP/PLAC-seq [168]	Jointly models the non-uniform coverage and genomic distance scaling of contact counts	FitHiChIP	76
[169]	2019	HiChIP/PLAC-seq	Zero-truncated Poisson regression framework	MAPS	65
[170]	2020	HiChIP	Differential peak analysis	HiChIP-Peaks	6
[171]	2010	ChIA-PET	Automatic processing of ChIA-PET data	ChIA-PET Tool	308
[172]	2014	ChIA-PET	A statistical model	chiasig	44
[173]	2015	ChIA-PET	R package to detect chromatin interactions from ChIA-PET	MICC	30
[174]	2015	ChIA-PET	Hierarchical Dirichlet process	3CPET	21
[175]	2017	ChIA-PET	Analysis pipeline	ChIA-PET2	71
[176]	2019	ChIA-PET	Analysis pipeline	ChIAPoP	5
[177]	2020	ChIA-PET	Analysis pipeline	ChIA-PIPE	8
[178]	2020	ChIA-PET	Consider different noise levels in different genomic regions	MACPET	0

ChIA-PIPE [177], and MACPET [178] have been developed for ChIA-PET data. Among these tools, HiC-Pro [158] is a pipeline tool for analyzing Hi-C data that includes data pre-processing and calling loops, and FitHiChIP [167] is a fast and memory-efficient loop caller for identifying significant loops. In addition, ChIA-PET2 [175] identifies loops in raw ChIA-PET sequencing reads of different types.

### Methods for identifying EPIs based on traditional machine-learning

The development of high-throughput sequencing technology has produced a huge amount of genomic information, relating to factors such as histone modification and chromatin accessibility. These factors data make it possible to recognize EPIs based on traditional machine learning methods. The basic idea is

to use different high-throughput genomic signals as input features of a traditional machine learning model to predict these interactions through statistical calculations. The TF and RNA polymerase ChIP-seq have been reported to be the factors data that can detect EPIs by analyzing epigenomic signals in enhancers and promoters, including TargetFinder [143], EPIP [154], and the XGBoost-based approach [179]. In recent years, boosting ensemble learning methods (*e.g.*, Adaboost [180], gradient boosting decision tree (GBDT) [181], and XGboost [182]) have been used to predict EPIs by constructing multiple weak classifiers. For example, Yu *et al.* [179] first generated EPI data sets based on chromatin contact data, annotated histones and binding protein data, and a GTF file, and then extracted epigenomic and sequence features. They then trained the XGBoost-based model by five-fold cross-validation in order to predict EPIs. They [179] showed that XGBoost

performed better than other machine learning methods, such as TargetFinder [143], random forest [147,183], GBDT [145], or Adaboost [154].

Methods based on traditional machine learning have the advantage of high accuracy for predicting EPIs. However, they have not been widely used for two reasons. The first is the lack of epigenetic characteristics in many cell lines, and the second is that traditional machine-learning-based methods require researchers to possess professional knowledge of epigenetics and manually connect the interaction characteristics.

### Methods for identifying EPIs based on deep-learning

With the development of deep learning, methods for identifying EPIs based on deep-learning have been proposed for building different neural network architectures in order to learn from DNA sequences without epigenomic characteristics. As is the case for the deep learning-based methods for predicting enhancers and promoters, the process of predicting EPIs includes three steps: (i) embedding the promoter and enhancer DNA sequences based on one-hot encoding or dna2vec, (ii) extracting the promoter and enhancer sequence features based on CNN, LSTM (long short-term memory), or transformer learning, and (iii) predicting EPIs based on the trained network.

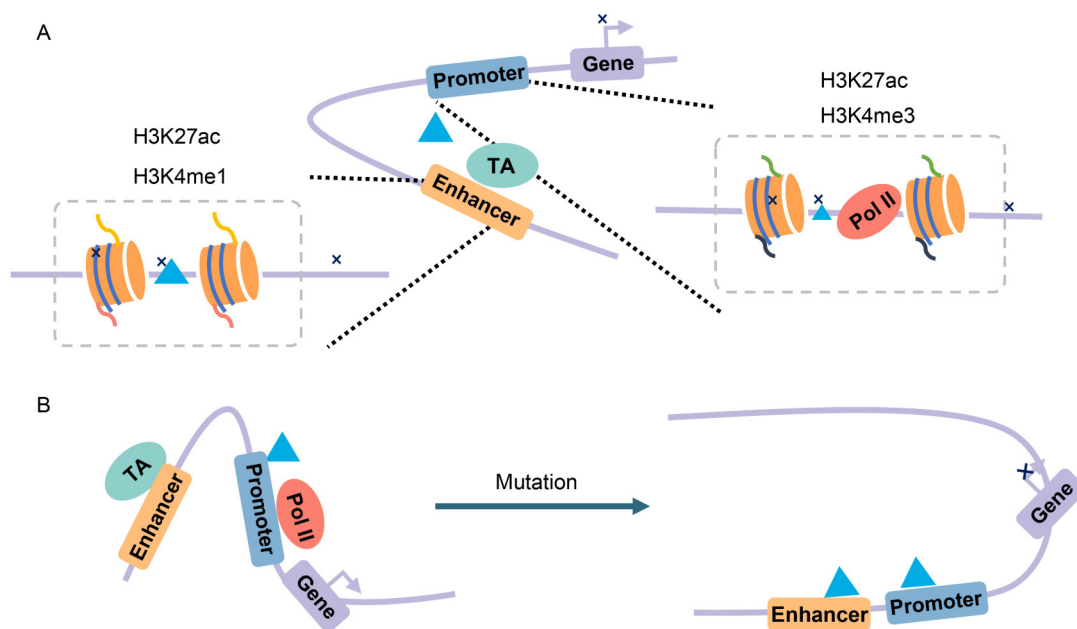
Zhuang *et al.* [148] used one-hot to encode the DNA sequence of enhancers and promoters, but the data storage needed for one-hot encoding consumes a great deal of computer memory and results in the loss of the association information among DNA sequences. EPIVAN [149] and EPI-Mind [153] use dna2vec to embed k-mer into a 100-dimensional vector and contained more information than was the case for one-hot encoding. Singh *et al.* [146] proposed SPEID to predict long-range EPIs that combine CNN with LSTM. SPEID [146] first inputs the enhancer and promoter vectors encoded by one-hot into CNN, fuses the high-dimensional features extracted from the enhancer and promoter, inputs the fused features into LSTM, and finally outputs the prediction results through the full connection layer. SEPT [155], EPIsHilbert [152], TransEPI [184], and EPI-Mind [153] used transfer learning to get more cross-cell type data features automatically. With the development of deep learning technology, applying transfer learning to the identification of EPIs can reduce the parameter training necessary for each different cell line.

Lastly, we counted the number of citations for available EPI tools, and found that TargetFinder [143] and IM-PET [42] were the most used EPI tools based on traditional machine-learning methods and that EPIVAN

[149] and SPEID [146] were the most used EPI tools based on deep-learning methods. Though the web server EPIXplore [185] has not been cited by any article, we suggest that users who do not want to run code access EPIXplore, because EPIXplore integrates IM-PET [42], EpiTensor [186], TargetFinder [143], JEME [187], and 3DPredictor [188], and provides downstream analysis as well as a visualization module. To explore the role that enhancer-promoter interaction structures play in determining normal and pathogenic cell states, we need to use tools that can identify differential EPIs in a process similar to differential expression analysis. Although there is no way to identify differential EPIs directly, we can combine the identification tools for differential loops and EPIs. For example, Lareau *et al.* proposed diffloop [189] to identify differential loops from ChIA-PET data and identified 1974 differential EPIs from 2 MCF7 and 2 K562 samples. diffHiC [190], FIND [191], HICcompare [192], multiHiCcompare [193], and Serpentine [194] all identify differential loops from Hi-C data.

### APPLICATIONS OF METHODS FOR IDENTIFYING EPIs IN DISEASES

Genome-wide association studies (GWAS) have revealed that noncoding regulatory sequences, especially the enhancer regions with strong cell specificity, are associated with disease variations [195,196]. Thus, any of the mutations that appear in enhancer-promoter interactions may cause diseases. Carullo *et al.* [197] discussed in their review study that two types of mutations may disrupt transcriptional regulation (Fig. 4A). First, the mutations of transcription factors or chromatin modifiers are found at enhancers. Marsman *et al.* [198] discussed the fact that the gene expression is regulated by transcription factors during cell development, and gene differentiation is regulated by changing loop conformations. For example, as Fig. 4B shows, the kit gene is expressed by transcription factors (*e.g.*, GATA-2) in immature erythrocytes, where the enhancers and kit promoter are linked via these transcription factors. When cells mature, other TFs (*e.g.*, GATA-1) that bind to the downstream element (DE) take the place of the GATA-2 TF. TFs including GATA-1 mediate looping between the kit promoter and DE, leading to the disappearance of the loop between enhancer and promoter and the downregulation of kit. Li *et al.* [199] also showed that the GATA-2 expression and DNA-binding are important for the cell differential process. Second, the mutations of sequence located in enhancers may lead to the loss or gain of functions. Wang *et al.* [200] proposed the model APRIL to construct long-range regulatory networks and predict



**Figure 4. Dynamic EPI affects gene transcription.** (A) Mutations at enhancers or promoters can lead to disease or to repressed gene expression. (B) The differential EPIs before and after cell mutation.

novel disease-associated genes with predicted enhancer-gene interactions as inputs (for example, from JEME [187] or IM-PET [42]). In a study by Rodin *et al.* [201], whole-genome sequencing was performed on 59 donors with autism spectrum disorder (ASD) and 15 control donors and functional enhancers provided by IM-PET [42] to demonstrate that ASD shows an excess of somatic mutations in neural enhancer sequences. Li *et al.* [18] suggested there is a possibility that mosaic enhancer mutations are associated with ASD risk. In addition, Fachal *et al.* [202] applied computational enhancer–promoter correlations (using IM-PET [42] and FANTOM5 [60]) and a Bayesian approach (PAINTOR) that they proposed to finely-map 150 breast cancer risk regions and identify 191 likely target genes.

## CONCLUSION AND FUTURE PROSPECTIVE

Computational methods for identification of enhancers, promoters, and EPIs are valuable for accelerating gene regulation studies, and this paper has reviewed the most important ones to come along over the past decade. We have proposed a basic framework for identifying EPIs and divided the identification methods of EPIs into the following two categories: (i) screening EPIs from ChIP-seq, Hi-C, HiChIP, ChIA-PET, or other High-throughput sequencing technology and (ii) identifying EPIs from DNA sequences, ChIP-seq, Hi-C, or other epigenome data by methods based on traditional machine learning or deep learning. This review also

covered enhancer and promoter databases (Table 3), as well as methods of identifying enhancers (Table 1), promoters (Table 2), chromatin loops (Table 5), and enhancer-promoter interactions (Table 6). These tables provide practical guidance for readers in selecting methods by model type or input data type in order to identify EPIs. We believe this review can serve as a foundational resource that allows researchers to apply traditional machine learning and deep learning methods to the prediction of enhancers, promoters, and EPIs in future research. We now summarize some important topics for this future work.

First, the initial step of EPI identification based on traditional machine-learning or deep-learning is to pre-process the DNA sequences using one-hot, k-mer, or dna2vec algorithms. However, these methods do not maintain the spatial proximity of the sequence. Designing a new sequence coding method that can maintain the spatial proximity and sequence features is the next task that we urge the EPI research community to undertake.

Secondly, although traditional machine-learning and deep-learning methods have furthered bioinformatics studies for enhancers, promoters, and EPIs for the past ten years, the precision of traditional machine learning is limited because of the high complexity of the source data, its features, and its limited possible model combinations. With recent increases in computing power, however, deep-learning-based methods for identifying EPIs directly from DNA sequences without other epigenome data features have begun to be

**Table 6 Prediction methods of enhancer-promoter interaction (EPI)**

Category	Refs.	Time	Source data	Method	Software name	Citation number
Traditional machine learning-based + call loops from Hi-C data	[42]	2014	DNA, histone marks, TFBSs, RNA-seq, ChIA-PET	Random forest	IM-PET	242
	[143]	2016	ChIP-seq, Hi-C	Machine learning-based	TargetFinder	349
	[154]	2019	Hi-C, enhancer, and promoter DNA sequences, ChIP-seq	Data screen, balanced and unbalanced models	EPIP	22
	[188]	2020	ChIP-seq, RNA-seq, Hi-C	Machine-learning-based	3DPredictor	32
Traditional machine learning-based	[203]	2017	ChIP-Seq	Bayesian classifier	EP_Bayes	8
	[187]	2017	DHS, distance, eRNA, histone marks, ChIA-PET/Hi-C/eQTL	Linear regression	JEME	166
	[183]	2017	5C, FAIRE-seq, ChIP-seq, Cap-analysis gene expression (CAGE), DNA methylation, nucleosome occupancy, eRNAs, chromatin state	Random forest classifier	–	11
	[144]	2017	DNA sequence	Gradient boosting	PEP	67
	[204]	2018	DNA structure properties and transcription factor binding motifs	Machine-learning-based	–	3
	[145]	2018	DNA sequences of arbitrary lengths	Natural language processing and unsupervised deep learning (extract sequence embedding feature), GBDT	EP2vec	56
	[147]	2019	ChIP-seq	Random forest	–	2
	[179]	2020	DNA sequence, ChIP-seq, annotation file	XGBoost-based	XGBoost	11
	[205]	2022	CT-FOCS	Linear mixed effect models	ct-focs	2
	Deep-learning-based	[148]	2019	DNA sequence	CNN and a recurrent neural network	EPIsCNN
[146]		2019	DNA sequence	CNN, LSTM	SPEID	94
[149]		2020	DNA sequence	Dna2vec, deep-learning-based	EPIVAN	100
[155]		2020	Hi-C, ChromHMM of Roadmap Epigenomics	CNN, transfer learning	SEPT	14
[151]		2021	DNA sequence	CNN, bidirectional gated recurrent unit network and matching heuristic mechanism	EPI-DLMH	18
[152]		2021	Hi-C, DNA sequence,	Hilbert curve encoding, transfer learning	EPIsHilbert	2
[184]		2022	Hi-C, ChIA-PET	Transformer-based model	TransEPI	1
[153]		2022	DNA sequence	Dna2vec, transfer learning	EPI-Mind	0
[185]	2022	–	A web server for prediction EPI	EPIXplorer	0	

developed. Furthermore, the rise of transfer learning has reduced the parameter training time needed by bioinformatics researchers. One model can even be fine-tuned by using transfer learning and then transferred to other models for training, which can significantly reduce the amount of needed calculations. For example, transfer learning can be used to predict EPIs [152,153,155,184] across cell lines. An appropriate model trained in one cell line can then be used to predict EPIs directly in another cell line, and this is something that we believe should become a research priority in the future.

Thirdly, with the development of single cell sequence technology, EPI studies at the single-cell level can help us solve the problem of cell heterogeneity, and analyze

the mechanism and relationship between individual cells and the body. To accomplish this, available EPI identification methods need to be optimized to accommodate the sparsity of single-cell sequencing data, such as scATAC-seq, scHi-C.

Fourthly, the application of EPI identification methods to exploring tumor-specific EPIs, the effect of mutations on EPIs, and the relationship between EPI formation and gene expression remains the central problem in EPI research. With the development of CRISPR technologies (CRISPR/Cas9, CRISPRa, CRISPRi) and CRISPR screening (Perturb-seq, CRISPRi-FlowFISH etc.), we are now able to identify EPIs or assess the role of EPIs in specific tumors and gene regulatory systems.

## ABBREVIATIONS

CREs	cis-acting regulatory elements
EPI(s)	Enhancer-promoter interaction(s)
TSS	Transcription start sites
ChIP-seq	Chromatin immunoprecipitation
CUT& RUN	Cleavage under targets and release using nuclease
Hi-C	High-throughput chromosome conformation capture
ChIA-PET	Chromatin interaction analysis with paired-end-tag sequencing
TFs	Transcription factors
TFBS	Transcription factor binding sites
CKSNAP	Composition of k-spaced nucleic acid pair
DCC	Dinucleotide-based cross covariance
PseDNC	Pseudo dinucleotide composition
PseKNC	Pseudo k-tuple nucleotide composition
SVM	Support vector machine
CNN	Convolution neural network
GBDT	Gradient boosting decision tree
LSTM	Long short-term memory
DE	Downstream element

## ACKNOWLEDGEMENTS

This study was funded by grants from the Foshan Higher Education Foundation (No. BKBS202203), the National Key R&D Program of China (No. 2018YFA0801402), the National Natural Science Foundation of China (No. 61971031) and the CAMS Innovation Fund for Medical Sciences (Nos. 2021-RC310-007, 2021-I2M-1-020 and 2022-I2M-1-020). Funding for open access charge: Department of Computer Science and Technology, Advanced Innovation Center for Materials Genome Engineering, University of Science and Technology Beijing. The authors thank AiMi Academic Services for English language editing and review services.

## COMPLIANCE WITH ETHICS GUIDELINES

Haiyan Gong, Zhengyuan Chen, Yuxin Tang, Minghong Li, Sichen Zhang, Xiaotong Zhang, and Yang Chen declare that they have no conflict of interest.

This article is a review article and does not contain any studies with human or animal subjects performed by any of the authors.

## OPEN ACCESS

This article is licensed by the CC BY under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated

otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## REFERENCES

- Bondarenko, V. A., Liu, Y. V., Jiang, Y. I. and Studitsky, V. M. (2003) Communication over a large distance: enhancers and insulators. *Biochem. Cell Biol.*, 81, 241–251
- Plank, J. L. and Dean, A. (2014) Enhancer function: mechanistic and genome-wide insights come together. *Mol. Cell*, 55, 5–14
- Haberle, V. and Lenhard, B. (2016) Promoter architectures and developmental gene regulation. *Semin. Cell Dev. Biol.*, 57, 11–23
- Harismendy, O., Notani, D., Song, X., Rahim, N. G., Tanasa, B., Heintzman, N., Ren, B., Fu, X. D., Topol, E. J., Rosenfeld, M. G., *et al.* (2011) 9p21 DNA variants associated with coronary artery disease impair interferon- $\gamma$  signalling response. *Nature*, 470, 264–268
- Luo, X., Liu, Y., Dang, D., Hu, T., Hou, Y., Meng, X., Zhang, F., Li, T., Wang, C., Li, M., *et al.* (2021) 3D Genome of macaque fetal brain reveals evolutionary innovations during primate corticogenesis. *Cell*, 184, 723–740.e21
- Schmitt, A. D., Hu, M., Jung, I., Xu, Z., Qiu, Y., Tan, C. L., Li, Y., Lin, S., Lin, Y., Barr, C. L., *et al.* (2016) A compendium of chromatin contact maps reveals spatially active regions in the human genome. *Cell Rep.*, 17, 2042–2059
- Policarpi, C., Crepaldi, L., Brookes, E., Nitaraska, J., French, S. M., Coatti, A. and Riccio, A. (2017) Enhancer sines link pol III to pol II transcription in neurons. *Cell Rep.*, 21, 2879–2894
- 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
- May, D., Blow, M. J., Kaplan, T., McCulley, D. J., Jensen, B. C., Akiyama, J. A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., *et al.* (2011) Large-scale discovery of enhancers from human heart tissue. *Nat. Genet.*, 44, 89–93
- Davison, L. J., Wallace, C., Cooper, J. D., Cope, N. F., Wilson, N. K., Smyth, D. J., Howson, J. M., Saleh, N., Al-Jeffery, A., Angus, K. L., *et al.* (2012) Long-range DNA looping and gene expression analyses identify DEXI as an autoimmune disease candidate gene. *Hum. Mol. Genet.*, 21, 322–333
- Smemo, S., Tena, J. J., Kim, K.-H., Gamazon, E. R., Sakabe, N. J., Gómez-Marín, C., Aneas, I., Credidio, F. L., Sobreira, D. R., Wasserman, N. F., *et al.* (2014) Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature*, 507, 371–375
- Schmidl, C., Rendeiro, A. F., Sheffield, N. C. and Bock, C. (2015) ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. *Nat. Methods*, 12, 963–965

13. Carey, M. F., Peterson, C. L., and Smale, S. T. (2009) Chromatin immunoprecipitation (ChIP). *Cold Spring Harb. Protoc.*, 2009, pdb.prot5279
14. Skene, P. J. and Henikoff, S. (2017) An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife*, 6, e21856
15. Belton, J. M., McCord, R. P., Gibcus, J. H., Naumova, N., Zhan, Y. and Dekker, J. (2012) Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods*, 58, 268–276
16. Liang, Z., Li, G., Wang, Z., Djekidel, M. N., Li, Y., Qian, M.-P., Zhang, M. Q. and Chen, Y. (2017) BL-Hi-C is an efficient and sensitive approach for capturing structural and regulatory chromatin interactions. *Nat. Commun.*, 8, 1622
17. Schoenfelder, S., Javierre, B. M., Furlan-Magaril, M., Wingett, S. W. and Fraser, P. (2018) Promoter capture Hi-C: high-resolution, genome-wide profiling of promoter interactions. *J. Vis. Exp.*, (136), e57320
18. 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
19. Mumbach, M. R., Rubin, A. J., Flynn, R. A., Dai, C., Khavari, P. A., Greenleaf, W. J. and Chang, H. Y. (2016) HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat. Methods*, 13, 919–922
20. Khan, Z. U., Pi, D. C., Yao, S. L., Nawaz, A., Ali, F. and Ali, S. (2021) Pienpred: a bi-layered discriminative model for enhancers and their subtypes via novel cascade multi-level subset feature selection algorithm. *Front. Comput. Sci.*, 15, 1–11
21. Visel, A., Rubin, E. M. and Pennacchio, L. A. (2009) Genomic views of distant-acting enhancers. *Nature*, 461, 199–205
22. Levine, M. (2010) Transcriptional enhancers in animal development and evolution. *Curr. Biol.*, 20, R754–R763
23. Bulger, M. and Groudine, M. (2011) Functional and mechanistic diversity of distal transcription enhancers. *Cell*, 144, 327–339
24. Kim, T. K. and Shiekhhattar, R. (2015) Architectural and functional commonalities between enhancers and promoters. *Cell*, 162, 948–959
25. Ong, C. T. and Corces, V. G. (2011) Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nat. Rev. Genet.*, 12, 283–293
26. Calo, E. and Wysocka, J. (2013) Modification of enhancer chromatin: what, how, and why? *Mol Cell*, 49, 825–837
27. Yao, L., Berman, B. P. and Farnham, P. J. (2015) Demystifying the secret mission of enhancers: linking distal regulatory elements to target genes. *Crit. Rev. Biochem. Mol. Biol.*, 50, 550–573
28. Klefogiannis, D., Kalnis, P. and Bajic, V. B. (2016) Progress and challenges in bioinformatics approaches for enhancer identification. *Brief. Bioinform.*, 17, 967–979
29. Lim, L. W. K., Chung, H. H., Chong, Y. L. and Lee, N. K. (2018) A survey of recently emerged genome-wide computational enhancer predictor tools. *Comput. Biol. Chem.*, 74, 132–141
30. Kaur, A., Chauhan, A. P. S., Aggarwal, A. K., and IEEE. (2019) Machine learning based comparative analysis of methods for enhancer prediction in genomic data. In: 2019 2nd International Conference on Intelligent Communication and Computational Techniques (ICCT), 142–145
31. Kyrchanova, O. and Georgiev, P. (2021) Mechanisms of enhancer-promoter interactions in higher eukaryotes. *Int. J. Mol. Sci.*, 22, 671
32. Mora, A., Sandve, G. K., Gabrielsen, O. S. and Eskeland, R. (2016) In the loop: promoter-enhancer interactions and bioinformatics. *Brief. Bioinform.*, 17, 980–995
33. Vanhaeren, T., Divina, F., García-Torres, M., Gómez-Vela, F., Vanhoof, W. and Martínez-García, P. M. (2020) A comparative study of supervised machine learning algorithms for the prediction of long-range chromatin interactions. *Genes (Basel)*, 11, 985
34. Tao, H., Li, H., Xu, K., Hong, H., Jiang, S., Du, G., Wang, J., Sun, Y., Huang, X., Ding, Y., *et al.* (2021) Computational methods for the prediction of chromatin interaction and organization using sequence and epigenomic profiles. *Brief. Bioinform.*, 22, bbaa405
35. Xu, H., Zhang, S., Yi, X., Plewczynski, D. and Li, M. J. (2020) Exploring 3D chromatin contacts in gene regulation: the evolution of approaches for the identification of functional enhancer-promoter interaction. *Comput. Struct. Biotechnol. J.*, 18, 558–570
36. He, C., Li, G., Nadhir, D. M., Chen, Y., Wang, X. and Zhang, M. Q. (2016) Advances in computational ChIA-PET data analysis. *Quant. Biol.*, 4, 217–225
37. Min, X., Lu, F. and Li, C. (2021) Sequence-based deep learning frameworks on enhancer-promoter interactions prediction. *Curr. Pharm. Des.*, 27, 1847–1855
38. Schoenfelder, S. and Fraser, P. (2019) Long-range enhancer-promoter contacts in gene expression control. *Nat. Rev. Genet.*, 20, 437–455
39. Kulaeva, O. I., Nizovtseva, E. V., Polikanov, Y. S., Ulianov, S. V. and Studitsky, V. M. (2012) Distant activation of transcription: mechanisms of enhancer action. *Mol. Cell Biol.*, 32, 4892–4897
40. Williams, A., Spilianakis, C. G. and Flavell, R. A. (2010) Interchromosomal association and gene regulation in trans. *Trends Genet.*, 26, 188–197
41. Maass, P. G., Barutcu, A. R. and Rinn, J. L. (2019) Interchromosomal interactions: a genomic love story of kissing chromosomes. *J. Cell Biol.*, 218, 27–38
42. He, B., Chen, C., Teng, L. and Tan, K. (2014) Global view of enhancer-promoter interactome in human cells. *Proc. Natl. Acad. Sci. USA*, 111, E2191–E2199
43. Patel, B., Kang, Y., Cui, K., Litt, M., Riberio, M. S., Deng, C., Salz, T., Casada, S., Fu, X., Qiu, Y., *et al.* (2014) Aberrant TAL1 activation is mediated by an interchromosomal interaction in human T-cell acute lymphoblastic leukemia. *Leukemia*, 28, 349–361

44. Lee, D., Karchin, R. and Beer, M. A. (2011) Discriminative prediction of mammalian enhancers from DNA sequence. *Genome Res.*, 21, 2167–2180
45. Taher, L., Narlikar, L. and Ovcharenko, I. (2012) Clare: cracking the language of regulatory elements. *Bioinformatics*, 28, 581–583
46. Fernández, M. and Miranda-Saavedra, D. (2012) Genome-wide enhancer prediction from epigenetic signatures using genetic algorithm-optimized support vector machines. *Nucleic Acids Res.*, 40, e77–e77
47. Rajagopal, N., Xie, W., Li, Y., Wagner, U., Wang, W., Stamatoyanopoulos, J., Ernst, J., Kellis, M. and Ren, B. (2013) RFECS: a random-forest based algorithm for enhancer identification from chromatin state. *PLOS Comput. Biol.*, 9, e1002968
48. Ghandi, M., Lee, D., Mohammad-Noori, M. and Beer, M. A. (2014) Enhanced regulatory sequence prediction using gapped k-mer features. *PLOS Comput. Biol.*, 10, e1003711
49. Erwin, G. D., Oksenberg, N., Truty, R. M., Kostka, D., Murphy, K. K., Ahituv, N., Pollard, K. S. and Capra, J. A. (2014) Integrating diverse datasets improves developmental enhancer prediction. *PLOS Comput. Biol.*, 10, e1003677
50. Lu, Y., Qu, W., Shan, G. and Zhang, C. (2015) Delta: a distal enhancer locating tool based on AdaBoost algorithm and shape features of chromatin modifications. *PLoS One*, 10, e0130622
51. Klefogiannis, D., Kalnis, P. and Bajic, V. B. (2015) DEEP: a general computational framework for predicting enhancers. *Nucleic Acids Res.*, 43, e6
52. Liu, B. (2016) Enhancer-pseudoknc: identification of enhancers and their subgroups based on pseudo degenerate kmer nucleotide composition. *Neurocomputing*, 217, 46–52
53. Liu, B., Fang, L., Long, R., Lan, X. and Chou, K. C. (2016) iEnhancer-2L: a two-layer predictor for identifying enhancers and their strength by pseudo k-tuple nucleotide composition. *Bioinformatics*, 32, 362–369
54. Jia, C. and He, W. (2016) EnhancerPred: a predictor for discovering enhancers based on the combination and selection of multiple features. *Sci. Rep.*, 6, 38741
55. Xu, J., Hu, H. and Dai, Y. (2016) Lmethyr-svm: predict human enhancers using low methylated regions based on weighted support vector machines. *PLoS One*, 11, e0163491
56. Colbran, L. L., Chen, L. and Capra, J. A. (2017) Short DNA sequence patterns accurately identify broadly active human enhancers. *BMC Genomics*, 18, 536
57. He, W. and Jia, C. (2017) EnhancerPred2.0: predicting enhancers and their strength based on position-specific trinucleotide propensity and electron-ion interaction potential feature selection. *Mol. Biosyst.*, 13, 767–774
58. He, Y., Gorkin, D. U., Dickel, D. E., Nery, J. R., Castanon, R. G., Lee, A. Y., Shen, Y., Visel, A., Pennacchio, L. A., Ren, B., *et al.* (2017) Improved regulatory element prediction based on tissue-specific local epigenomic signatures. *Proc. Natl. Acad. Sci. USA*, 114, E1633–E1640
59. Liu, B., Li, K., Huang, D. S. and Chou, K. C. (2018) iEnhancer-EL: identifying enhancers and their strength with ensemble learning approach. *Bioinformatics*, 34, 3835–3842
60. Klefogiannis, D., Ashoor, H. and Bajic, V. B. (2018) Tels: a novel computational framework for identifying motif signatures of transcribed enhancers. *Genom. Proteom. Bioinf.*, 16, 332–341
61. Singh, A. P., Mishra, S. and Jabin, S. (2018) Sequence based prediction of enhancer regions from DNA random walk. *Sci. Rep.*, 8, 15912
62. Sethi, A., Gu, M., Gumusgoz, E., Chan, L., Yan, K. K., Rozowsky, J., Barozzi, I., Afzal, V., Akiyama, J. A., Plajzer-Frick, I., *et al.* (2020) Supervised enhancer prediction with epigenetic pattern recognition and targeted validation. *Nat. Methods*, 17, 807–814
63. Lim, D. Y., Khanal, J., Tayara, H. and Chong, K. T. (2021) Enhancer-*rf*: identifying enhancers and their strength by enhanced feature representation using random forest. *Chemom. Intell. Lab. Syst.*, 212, 104284
64. Lyu, Y., Zhang, Z., Li, J., He, W., Ding, Y. and Guo, F. (2021) Enhancer-*kl*: a novel two-layer predictor for identifying enhancers by position specific of nucleotide composition. *IEEE/ACM Trans. Comput. Biol. Bioinformatics*, 18, 2809–2815
65. Niu, X., Deng, K., Liu, L., Yang, K. and Hu, X. (2021) A statistical framework for predicting critical regions of p53-dependent enhancers. *Brief. Bioinform.*, 22, bbaa053
66. Basith, S., Hasan, M. M., Lee, G., Wei, L. and Manavalan, B. (2021) Integrative machine learning framework for the identification of cell-specific enhancers from the human genome. *Brief. Bioinform.*, 22, bbab252
67. Firpi, H. A., Ucar, D. and Tan, K. (2010) Discover regulatory DNA elements using chromatin signatures and artificial neural network. *Bioinformatics*, 26, 1579–1586
68. Liu, F., Li, H., Ren, C., Bo, X. and Shu, W. (2016) PEDLA: predicting enhancers with a deep learning-based algorithmic framework. *Sci. Rep.*, 6, 28517
69. Min, X., Zeng, W., Chen, S., Chen, N., Chen, T. and Jiang, R. (2017) Predicting enhancers with deep convolutional neural networks. *BMC Bioinformatics*, 18, 478
70. Yang, B., Liu, F., Ren, C., Ouyang, Z., Xie, Z., Bo, X. and Shu, W. (2017) BiRen: predicting enhancers with a deep-learning-based model using the DNA sequence alone. *Bioinformatics*, 33, 1930–1936
71. Thibodeau, A., Uyar, A., Khetan, S., Stitzel, M. L. and Ucar, D. (2018) A neural network based model effectively predicts enhancers from clinical ATAC-seq samples. *Sci. Rep.*, 8, 16048
72. Le, N. Q. K., Yapp, E. K. Y., Ho, Q. T., Nagasundaram, N., Ou, Y. Y. and Yeh, H. Y. (2019) iEnhancer-5Step: identifying enhancers using hidden information of DNA sequences via Chou's 5-step rule and word embedding. *Anal. Biochem.*, 571, 53–61
73. Khanal, J., Tayara, H. and Chong, K. T. (2020) Identifying enhancers and their strength by the integration of word embedding and convolution neural network. *IEEE Access*, 8,

- 58369–58376
74. Chen, S., Gan, M., Lv, H. and Jiang, R. (2021) Deepcape: a deep convolutional neural network for the accurate prediction of enhancers. *Genom. Proteom. Bioinf.*, 19, 565–577
  75. Chen, Z., Zhang, J., Liu, J., Dai, Y., Lee, D., Min, M. R., Xu, M. and Gerstein, M. (2021) DECODE: a Deep-learning framework for condensing enhancers and refining boundaries with large-scale functional assays. *Bioinformatics*, 37, i280–i288
  76. Zhang, T. H., Flores, M. and Huang, Y. (2021) ES-ARCNN: predicting enhancer strength by using data augmentation and residual convolutional neural network. *Anal. Biochem.*, 618, 114120
  77. Inayat, N., Khan, M., Iqbal, N., Khan, S., Raza, M., Khan, D. M., Khan, A. and Wei, D. Q. (2021) Ienhancer-dhf: identification of enhancers and their strengths using optimize deep neural network with multiple features extraction methods. *IEEE Access*, 9, 40783–40796
  78. Yang, R., Wu, F., Zhang, C. and Zhang, L. (2021) Ienhancer-gan: a deep learning framework in combination with word embedding and sequence generative adversarial net to identify enhancers and their strength. *Int. J. Mol. Sci.*, 22, 3589
  79. Gao, Y., Chen, Y., Feng, H., Zhang, Y. and Yue, Z. (2022) Ricenn: prediction of rice enhancers with neural network based on DNA sequences. *Interdiscip. Sci.*, 14, 555–565
  80. Amilpur, S. and Bhukya, R. (2022) A sequence-based two-layer predictor for identifying enhancers and their strength through enhanced feature extraction. *J. Bioinform. Comput. Biol.*, 20, 2250005
  81. Kamran, H., Tahir, M., Tayara, H. and Chong, K. T. (2022) Ienhancer-deep: a computational predictor for enhancer sites and their strength using deep learning. *Appl. Sci. (Basel)*, 12, 2120
  82. Zhao, S., Pan, Q., Zou, Q., Ju, Y., Shi, L. and Su, X. (2022) Identifying and classifying enhancers by dinucleotide-based auto-cross covariance and attention-based Bi-LSTM. *Comput. Math. Methods Med.*, 2022, 7518779
  83. Song, K. (2012) Recognition of prokaryotic promoters based on a novel variable-window Z-curve method. *Nucleic Acids Res.*, 40, 963–971
  84. Li, Y., Chen, C. Y. and Wasserman, W. W. (2016) Deep feature selection: theory and application to identify enhancers and promoters. *J. Comput. Biol.*, 23, 322–336
  85. Umarov, R. K. and Solovyev, V. V. (2017) Recognition of prokaryotic and eukaryotic promoters using convolutional deep learning neural networks. *PLoS One*, 12, e0171410
  86. Coelho, R. V., de Avila E Silva, S., Echeverrigaray, S., Delamare, A. P. L. and Delamare, A. P. L. (2018) *Bacillus subtilis* promoter sequences data set for promoter prediction in Gram-positive bacteria. *Data Brief*, 19, 264–270
  87. Liu, B., Yang, F., Huang, D. S. and Chou, K. C. (2018) iPromoter-2L: a two-layer predictor for identifying promoters and their types by multi-window-based PseKNC. *Bioinformatics*, 34, 33–40
  88. Oubounyt, M., Louadi, Z., Tayara, H. and Chong, K. T. (2019) Deepromoter: robust promoter predictor using deep learning. *Front. Genet.*, 10, 286
  89. Le, N. Q. K., Yapp, E. K. Y., Nagasundaram, N. and Yeh, H. Y. (2019) Classifying promoters by interpreting the hidden information of DNA sequences via deep learning and combination of continuous fasttext N-grams. *Front. Bioeng. Biotechnol.*, 7, 305
  90. Umarov, R., Kuwahara, H., Li, Y., Gao, X. and Solovyev, V. (2019) Promoter analysis and prediction in the human genome using sequence-based deep learning models. *Bioinformatics*, 35, 2730–2737
  91. Lai, H. Y., Zhang, Z. Y., Su, Z. D., Su, W., Ding, H., Chen, W. and Lin, H. (2019) Iproep: a computational predictor for predicting promoter. *Mol. Ther. Nucleic Acids*, 17, 337–346
  92. Liu, B. and Li, K. (2019) Ipromoter-2l2: identifying promoters and their types by combining smoothing cutting window algorithm and sequence-based features. *Mol. Ther. Nucleic Acids*, 18, 80–87
  93. Rahman, M. S., Aktar, U., Jani, M. R. and Shatabda, S. (2019) iPromoter-FSEn: identification of bacterial  $\sigma^{70}$  promoter sequences using feature subspace based ensemble classifier. *Genomics*, 111, 1160–1166
  94. Rahman, M. S., Aktar, U., Jani, M. R. and Shatabda, S. (2019) iPro70-FMWin: identifying Sigma70 promoters using multiple windowing and minimal features. *Mol. Genet. Genomics*, 294, 69–84
  95. Xiao, X., Xu, Z. C., Qiu, W. R., Wang, P., Ge, H. T. and Chou, K. C. (2019) iPSW(2L)-PseKNC: a two-layer predictor for identifying promoters and their strength by hybrid features via pseudo k-tuple nucleotide composition. *Genomics*, 111, 1785–1793
  96. Zhang, M., Li, F., Marquez-Lago, T. T., Leier, A., Fan, C., Kwok, C. K., Chou, K. C., Song, J. and Jia, C. (2019) MULTiPly: a novel multi-layer predictor for discovering general and specific types of promoters. *Bioinformatics*, 35, 2957–2965
  97. Chen, Y. L., Guo, D. H. and Li, Q. Z. (2020) An energy model for recognizing the prokaryotic promoters based on molecular structure. *Genomics*, 112, 2072–2079
  98. Amin, R., Rahman, C. R., Ahmed, S., Sifat, M. H. R., Liton, M. N. K., Rahman, M. M., Khan, M. Z. H. and Shatabda, S. (2020) iPromoter-BnCNN: a novel branched CNN-based predictor for identifying and classifying sigma promoters. *Bioinformatics*, 36, 4869–4875
  99. Tayara, H., Tahir, M. and Chong, K. T. (2020) Identification of prokaryotic promoters and their strength by integrating heterogeneous features. *Genomics*, 112, 1396–1403
  100. Shujaat, M., Wahab, A., Tayara, H. and Chong, K. T. (2020) Pcpromoter-CNN: a CNN-based prediction and classification of promoters. *Genes (Basel)*, 11, 1529
  101. Liang, Y., Zhang, S., Qiao, H. and Yao, Y. (2021) iPromoter-ET: identifying promoters and their strength by extremely randomized trees-based feature selection. *Anal. Biochem.*, 630,

- 114335
102. Shujaat, M., Lee, S. B., Tayara, H. and Chong, K. T. (2021) Cr-prom: a convolutional neural network-based model for the prediction of rice promoters. *IEEE Access*, 9, 81485–81491
  103. Lyu, Y., He, W., Li, S., Zou, Q. and Guo, F. (2021) Ipro2l-pstknc: a two-layer predictor for discovering various types of promoters by position specific of nucleotide composition. *IEEE J. Biomed. Health Inform.*, 25, 2329–2337
  104. Sun, A., Xiao, X. and Xu, Z. (2021) Iptt(2 L)-CNN: a two-layer predictor for identifying promoters and their types in plant genomes by convolutional neural network. *Comput. Math. Methods Med.*, 2021, 6636350
  105. Zhu, Y., Li, F., Xiang, D., Akutsu, T., Song, J. and Jia, C. (2021) Computational identification of eukaryotic promoters based on cascaded deep capsule neural networks. *Brief. Bioinform.*, 22, bbaa299
  106. Bhukya, R., Kumari, A., Amilpur, S. and Dasari, C. M. (2022) PPred-PCKSM: a multi-layer predictor for identifying promoter and its variants using position based features. *Comput. Biol. Chem.*, 97, 107623
  107. Li, H., Shi, L., Gao, W., Zhang, Z., Zhang, L., Zhao, Y. and Wang, G. (2022) dPromoter-XGBoost: detecting promoters and strength by combining multiple descriptors and feature selection using XGBoost. *Methods*, 204, 215–222
  108. Li, Q. W., Zhang, L. C., Xu, L., Zou, Q., Wu, J. and Li, Q. Y. (2022) Identification and classification of promoters using the attention mechanism based on long short-term memory. *Front. Comput. Sci.*, 16, 164348
  109. Qiao, H., Zhang, S., Xue, T., Wang, J. and Wang, B. (2022) iPro-GAN: a novel model based on generative adversarial learning for identifying promoters and their strength. *Comput. Methods Programs Biomed.*, 215, 106625
  110. Wang, Y., Peng, Q., Mou, X., Wang, X., Li, H., Han, T., Sun, Z. and Wang, X. (2022) A successful hybrid deep learning model aiming at promoter identification. *BMC Bioinformatics*, 23, 206
  111. Wei, P. J., Pang, Z. Z., Jiang, L. J., Tan, D. Y., Su, Y. S. and Zheng, C. H. (2022) Promoter prediction in nannochloropsis based on densely connected convolutional neural networks. *Methods*, 204, 38–46
  112. 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
  113. Javierre, B. M., Burren, O. S., Wilder, S. P., Kreuzhuber, R., Hill, S. M., Sewitz, S., Cairns, J., Wingett, S. W., Várnai, 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.e19
  114. Mikolov, T., Yih, W.-t. and Zweig, G. (2013) Linguistic regularities in continuous space word representations. In: *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human language technologies*, 746–751
  115. Pennington, J., Socher, R. and Manning, C. D. (2014) Glove: global vectors for word representation. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543
  116. Ng, P. (2017) Dna2vec: consistent vector representations of variable-length k-mers. *arXiv*, 1701.06279
  117. Jiang, Y., Qian, F., Bai, X., Liu, Y., Wang, Q., Ai, B., Han, X., Shi, S., Zhang, J., Li, X., *et al.* (2019) SEdb: a comprehensive human super-enhancer database. *Nucleic Acids Res.*, 47, D235–D243
  118. Périer, R. C., Praz, V., Junier, T., Bonnard, C. and Bucher, P. (2000) The eukaryotic promoter database (EPD). *Nucleic Acids Res.*, 28, 302–303
  119. Ferretti, V., Poitras, C., Bergeron, D., Coulombe, B., Robert, F. and Blanchette, M. (2007) PReMod: a database of genome-wide mammalian cis-regulatory module predictions. *Nucleic Acids Res.*, 35, D122–D126
  120. Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmid, C., Suzuki, T., *et al.* (2014) An atlas of active enhancers across human cell types and tissues. *Nature*, 507, 455–461
  121. Visel, A., Minovitsky, S., Dubchak, I. and Pennacchio, L. A. (2007) VISTA Enhancer Browser—a database of tissue-specific human enhancers. *Nucleic Acids Res.*, 35, D88–D92
  122. Lovén, J., Hoke, H. A., Lin, C. Y., Lau, A., Orlando, D. A., Vakoc, C. R., Bradner, J. E., Lee, T. I. and Young, R. A. (2013) Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell*, 153, 320–334
  123. Bai, X., Shi, S., Ai, B., Jiang, Y., Liu, Y., Han, X., Xu, M., Pan, Q., Wang, F., Wang, Q., *et al.* (2020) ENdb: a manually curated database of experimentally supported enhancers for human and mouse. *Nucleic Acids Res.*, 48, D51–D57
  124. Wei, Y., Zhang, S., Shang, S., Zhang, B., Li, S., Wang, X., Wang, F., Su, J., Wu, Q., Liu, H., *et al.* (2016) SEA: a super-enhancer archive. *Nucleic Acids Res.*, 44, D172–D179
  125. Cai, Z. N., Cui, Y., Tan, Z. Y., Zhang, G. H., Tan, Z. Y., Zhang, X. L. and Peng, Y. S. (2019) RAEdB: A database of enhancers identified by high-throughput reporter assays. *Database (Oxford)*, bay140
  126. Guo, Z. W., Xie, C., Li, K., Zhai, X. M., Cai, G. X., Yang, X. X. and Wu, Y. S. (2019) Seler: a database of super-enhancer-associated lncRNA-directed transcriptional regulation in human cancers. *Database (Oxford)*, baz027
  127. Zeng, W. W., Min, X. and Jiang, R. (2019) Endisease: a manually curated database for enhancer-disease associations. *Database (Oxford)*, baz020
  128. Huang, M., Wang, Y., Yang, M., Yan, J., Yang, H., Zhuang, W., Xu, Y., Koeffler, H. P., Lin, D. C. and Chen, X. (2020) dbInDel: a database of enhancer-associated insertion and deletion variants by analysis of H3K27ac ChIP-Seq. *Bioinformatics*, 36, 1649–1651
  129. Kumar, R., Lathwal, A., Kumar, V., Patiyal, S., Raghav, P. K. and Raghava, G. P. S. (2020) CancerEnD: a database of cancer associated enhancers. *Genomics*, 112, 3696–3702

130. Vasyuchenko, E. P., Orekhov, P. S., Armeev, G. A. and Bozdaganyan, M. E. (2021) Cpe-db: an open database of chemical penetration enhancers. *Pharmaceutics*, 13, 66
131. Jin, W., Jiang, G., Yang, Y., Yang, J., Yang, W., Wang, D., Niu, X., Zhong, R., Zhang, Z. and Gong, J. (2022) Animal-eRNAdb: a comprehensive animal enhancer RNA database. *Nucleic Acids Res.*, 50, D46–D53
132. Shahmuradov, I. A., Gammerman, A. J., Hancock, J. M., Bramley, P. M. and Solovyev, V. V. (2003) PlantProm: a database of plant promoter sequences. *Nucleic Acids Res.*, 31, 114–117
133. Smirnova, O. G., Ibragimova, S. S. and Kochetov, A. V. (2012) Simple database to select promoters for plant transgenesis. *Transgenic Res.*, 21, 429–437
134. Grienberg, I. and Benayahu, D. (2005) Osteo-Promoter Database (OPD)—promoter analysis in skeletal cells. *BMC Genomics*, 6, 46
135. Morris, R. T., O'Connor, T. R. and Wyrick, J. J. (2008) Osiris: an integrated promoter database for *Oryza sativa* L. *Bioinformatics*, 24, 2915–2917
136. Chen, X., Wu, J. M., Hornischer, K., Kel, A. and Wingender, E. (2006) Tiprod: the tissue-specific promoter database. *Nucleic Acids Res.*, 34, D104–D107
137. Nishikata, K., Cox, R. S. 3rd, Shimoyama, S., Yoshida, Y., Matsui, M., Makita, Y. and Toyoda, T. (2014) Database construction for PromoterCAD: synthetic promoter design for mammals and plants. *ACS Synth. Biol.*, 3, 192–196
138. Dreos, R., Ambrosini, G., Périer, R. C. and Bucher, P. (2015) The Eukaryotic Promoter Database: expansion of EPDnew and new promoter analysis tools. *Nucleic Acids Res.*, 43, D92–D96
139. Su, W., Liu, M. L., Yang, Y. H., Wang, J. S., Li, S. H., Lv, H., Dao, F. Y., Yang, H. and Lin, H. (2021) Ppd: a manually curated database for experimentally verified prokaryotic promoters. *J. Mol. Biol.*, 433, 166860
140. Gordon, L., Chervonenkis, A. Y., Gammerman, A. J., Shahmuradov, I. A. and Solovyev, V. V. (2003) Sequence alignment kernel for recognition of promoter regions. *Bioinformatics*, 19, 1964–1971
141. Towsey, M., Timms, P., Hogan, J. and Mathews, S. A. (2008) The cross-species prediction of bacterial promoters using a support vector machine. *Comput. Biol. Chem.*, 32, 359–366
142. Knudsen, S. (1999) Promoter2.0: for the recognition of PolII promoter sequences. *Bioinformatics*, 15, 356–361
143. 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
144. Yang, Y., Zhang, R., Singh, S. and Ma, J. (2017) Exploiting sequence-based features for predicting enhancer-promoter interactions. *Bioinformatics*, 33, i252–i260
145. Zeng, W., Wu, M. and Jiang, R. (2018) Prediction of enhancer-promoter interactions via natural language processing. *BMC Genomics*, 19, 84
146. Singh, S., Yang, Y., Póczos, B. and Ma, J. (2019) Predicting enhancer-promoter interaction from genomic sequence with deep neural networks. *Quant. Biol.*, 7, 122–137
147. Zhang, T. and Wang, Y. (2019) An approach for recognition of enhancer-promoter associations based on random forest. In: *Proceedings of the 2019 4th International Conference on Biomedical Signal and Image Processing (ICBIP 2019)*, 46–50
148. Zhuang, Z., Shen, X. and Pan, W. (2019) A simple convolutional neural network for prediction of enhancer-promoter interactions with DNA sequence data. *Bioinformatics*, 35, 2899–2906
149. Hong, Z., Zeng, X., Wei, L. and Liu, X. (2020) Identifying enhancer-promoter interactions with neural network based on pre-trained DNA vectors and attention mechanism. *Bioinformatics*, 36, 1037–1043
150. Singh, A. P., Mishra, S., and Jabin, S. (2018) Sequence based prediction of enhancer regions from DNA random walk. *Sci Rep.* 8, 15912
151. Min, X., Ye, C., Liu, X. and Zeng, X. (2021) Predicting enhancer-promoter interactions by deep learning and matching heuristic. *Brief. Bioinform.*, 22, bbaa254
152. Zhang, M., Hu, Y. and Zhu, M. (2021) Epishilbert: prediction of enhancer-promoter interactions via hilbert curve encoding and transfer learning. *Genes (Basel)*, 12, 1385
153. Ni, Y., Fan, L., Wang, M., Zhang, N., Zuo, Y. and Liao, M. (2022) Epi-mind: identifying enhancer-promoter interactions based on transformer mechanism. *Interdiscip. Sci.*, 14, 786–794
154. Talukder, A., Saadat, S., Li, X. and Hu, H. (2019) EPIP: a novel approach for condition-specific enhancer-promoter interaction prediction. *Bioinformatics*, 35, 3877–3883
155. Jing, F., Zhang, S. W. and Zhang, S. (2020) Prediction of enhancer-promoter interactions using the cross-cell type information and domain adversarial neural network. *BMC Bioinformatics*, 21, 507
156. Chawla, N. V., Bowyer, K. W., Hall, L. O. and Kegelmeyer, W. P. (2002) Smote: synthetic minority over-sampling technique. *J. Artif. Intell. Res.*, 16, 321–357
157. Rao, S. S. P., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., Sanborn, A. L., Machol, I., Omer, A. D., Lander, E. S., *et al.* (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, 159, 1665–1680
158. Servant, N., Varoquaux, N., Lajoie, B. R., Viara, E., Chen, C.-J., Vert, J.-P., Heard, E., Dekker, J. and Barillot, E. (2015) HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.*, 16, 259
159. Kerpedjiev, P., Abdennur, N., Lekschas, F., McCallum, C., Dinkla, K., Strobelt, H., Lubner, J. M., Ouellette, S. B., Azhir, A., Kumar, N., *et al.* (2018) HiGlass: web-based visual exploration and analysis of genome interaction maps. *Genome Biol.*, 19, 125
160. Cao, Y., Chen, Z., Chen, X., Ai, D., Chen, G., McDermott, J., Huang, Y., Guo, X. and Han, J. J. (2020) Accurate loop calling for 3D genomic data with cLoops. *Bioinformatics*, 36, 666–675
161. Kaul, A., Bhattacharyya, S. and Ay, F. (2020) Identifying statistically significant chromatin contacts from Hi-C data with

- FitHiC2. *Nat. Protoc.*, 15, 991–1012
162. Roayaei Ardakany, A., Gezer, H. T., Lonardi, S. and Ay, F. (2020) Mustache: multi-scale detection of chromatin loops from Hi-C and Micro-C maps using scale-space representation. *Genome Biol.*, 21, 256
  163. Krietenstein, N., Abraham, S., Venev, S. V., Abdennur, N., Gibcus, J., Hsieh, T. S., Parsi, K. M., Yang, L., Maehr, R., Mirny, L. A., *et al.* (2020) Ultrastructural details of mammalian chromosome architecture. *Mol. Cell*, 78, 554–565.e7
  164. Lagler, T. M., Abnoui, A., Hu, M., Yang, Y. and Li, Y. (2021) HiC-ACT: improved detection of chromatin interactions from Hi-C data via aggregated Cauchy test. *Am. J. Hum. Genet.*, 108, 257–268
  165. Lee, H. and Seo, P. J. (2021) Hicore: Hi-c analysis for identification of core chromatin looping regions with higher resolution. *Mol. Cells*, 44, 883–892
  166. Lareau, C. A. and Aryee, M. J. (2018) hichipper: a preprocessing pipeline for calling DNA loops from HiChIP data. *Nat. Methods*, 15, 155–156
  167. Bhattacharyya, S., Chandra, V., Vijayanand, P. and Ay, F. (2019) Identification of significant chromatin contacts from HiChIP data by FitHiChIP. *Nat. Commun.*, 10, 4221
  168. Fang, R., Yu, M., Li, G., Chee, S., Liu, T., Schmitt, A. D. and Ren, B. (2016) Mapping of long-range chromatin interactions by proximity ligation-assisted ChIP-seq. *Cell Res.*, 26, 1345–1348
  169. Juric, I., Yu, M., Abnoui, A., Raviram, R., Fang, R., Zhao, Y., Zhang, Y., Qiu, Y., Yang, Y., Li, Y., *et al.* (2019) MAPS: model-based analysis of long-range chromatin interactions from PLAC-seq and HiChIP experiments. *PLOS Comput. Biol.*, 15, e1006982
  170. Shi, C., Rattray, M. and Orozco, G. (2020) HiChIP-Peaks: a HiChIP peak calling algorithm. *Bioinformatics*, 36, 3625–3631
  171. Li, G., Fullwood, M. J., Xu, H., Mulawadi, F. H., Velkov, S., Vega, V., Ariyaratne, P. N., Mohamed, Y. B., Ooi, H. S., Tennakoon, C., *et al.* (2010) ChIA-PET tool for comprehensive chromatin interaction analysis with paired-end tag sequencing. *Genome Biol.*, 11, R22
  172. Paulsen, J., Rødland, E. A., Holden, L., Holden, M. and Hovig, E. (2014) A statistical model of ChIA-PET data for accurate detection of chromatin 3D interactions. *Nucleic Acids Res.*, 42, e143
  173. He, C., Zhang, M. Q. and Wang, X. (2015) MICC: an R package for identifying chromatin interactions from ChIA-PET data. *Bioinformatics*, 31, 3832–3834
  174. 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
  175. Li, G., Chen, Y., Snyder, M. P. and Zhang, M. Q. (2017) ChIA-PET2: a versatile and flexible pipeline for ChIA-PET data analysis. *Nucleic Acids Res.*, 45, e4
  176. Huang, W., Medvedovic, M., Zhang, J. and Niu, L. (2019) ChIA-PoP: a new tool for ChIA-PET data analysis. *Nucleic Acids Res.*, 47, e37
  177. Lee, B., Wang, J., Cai, L., Kim, M., Namburi, S., Tjong, H., Feng, Y., Wang, P., Tang, Z., Abbas, A., *et al.* (2020) ChIA-PIPE: a fully automated pipeline for comprehensive ChIA-PET data analysis and visualization. *Sci. Adv.*, 6, eaay2078
  178. Vardaxis, I., Drabløs, F., Rye, M. B. and Lindqvist, B. H. (2020) MACPET: model-based analysis for ChIA-PET. *Biostatistics*, 21, 625–639
  179. Yu, X. J., Zhou, J. G., Zhao, M. M., Yi, C., Duan, Q., Zhou, W. and Li, J. (2020) Exploiting XGboost for predicting enhancer-promoter interactions. *Curr. Bioinform.*, 15, 1036–1045
  180. Bartlett, P. and Traskin, M. (2006) Adaboost is consistent. *Adv. Neural Inf. Process. Syst.*, 8, 2347–2368
  181. Zhang, S., Dong, X. and Guan, Y. (2012) Synonym recognition based on user behaviors in E-commerce. *Journal of Chinese Information Processing (in Chinese)*, 26, 79–85
  182. Chen, T. and Guestrin, C. (2016) XgBoost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794
  183. Feng, Z. X. and Li, Q. Z. (2017) Recognition of long-range enhancer-promoter interactions by adding genomic signatures of segmented regulatory regions. *Genomics*, 109, 341–352
  184. Chen, K., Zhao, H. and Yang, Y. (2022) Capturing large genomic contexts for accurately predicting enhancer-promoter interactions. *Brief. Bioinform.*, 23, bbab577
  185. Tang, L., Zhong, Z., Lin, Y., Yang, Y., Wang, J., Martin, J. F. and Li, M. (2022) EPIXplorer: a web server for prediction, analysis and visualization of enhancer-promoter interactions. *Nucleic Acids Res.*, 50, W290–W297
  186. Zhu, Y., Chen, Z., Zhang, K., Wang, M., Medovoy, D., Whitaker, J. W., Ding, B., Li, N., Zheng, L. and Wang, W. (2016) Constructing 3D interaction maps from 1D epigenomes. *Nat. Commun.*, 7, 10812
  187. 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
  188. Belokopytova, P. S., Nuriddinov, M. A., Mozheiko, E. A., Fishman, D. and Fishman, V. (2020) Quantitative prediction of enhancer-promoter interactions. *Genome Res.*, 30, 72–84
  189. Lareau, C. A. and Aryee, M. J. (2018) diffloop: a computational framework for identifying and analyzing differential DNA loops from sequencing data. *Bioinformatics*, 34, 672–674
  190. Lun, A. T. and Smyth, G. K. (2015) diffHic: a Bioconductor package to detect differential genomic interactions in Hi-C data. *BMC Bioinformatics*, 16, 258
  191. Djekidel, M. N., Chen, Y. and Zhang, M. Q. (2018) FIND: differential chromatin INteractions Detection using a spatial Poisson process. *Genome Res.*, 28, 412–422
  192. Stansfield, J. C., Cresswell, K. G., Vladimirov, V. I. and Dozmorov, M. G. (2018) HiCcompare: an R-package for joint normalization and comparison of Hi-C datasets. *BMC*

- Bioinformatics, 19, 279
193. Stansfield, J. C., Cresswell, K. G. and Dozmorov, M. G. (2019) multiHiCcompare: joint normalization and comparative analysis of complex Hi-C experiments. *Bioinformatics*, 35, 2916–2923
  194. Baudry, L., Millot, G. A., Thierry, A., Koszul, R. and Scolari, V. F. (2020) Serpentine: a flexible 2D binning method for differential Hi-C analysis. *Bioinformatics*, 36, 3645–3651
  195. Karczewski, K. J., Dudley, J. T., Kukurba, K. R., Chen, R., Butte, A. J., Montgomery, S. B. and Snyder, M. (2013) Systematic functional regulatory assessment of disease-associated variants. *Proc. Natl. Acad. Sci. USA*, 110, 9607–9612
  196. Corradin, O., Saiakhova, A., Akhtar-Zaidi, B., Myeroff, L., Willis, J., Cowper-Salari, R., Lupien, M., Markowitz, S. and Scacheri, P. C. (2014) Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits. *Genome Res.*, 24, 1–13
  197. Carullo, N. V. N. and Day, J. J. (2019) Genomic enhancers in brain health and disease. *Genes (Basel)*, 10, 43
  198. Marsman, J. and Horsfield, J. A. (2012) Long distance relationships: enhancer-promoter communication and dynamic gene transcription. *Gene Regulatory Mechanisms.*, 1819, 1217–1227
  199. Li, Y., He, Y., Liang, Z., Wang, Y., Chen, F., Djekidel, M. N., Li, G., Zhang, X., Xiang, S., Wang, Z., *et al.* (2018) Alterations of specific chromatin conformation affect ATRA-induced leukemia cell differentiation. *Cell Death Dis.*, 9, 200
  200. Wang, H., Yang, J., Zhang, Y. and Wang, J. (2021) Discover novel disease-associated genes based on regulatory networks of long-range chromatin interactions. *Methods*, 189, 22–33
  201. Rodin, R. E., Dou, Y., Kwon, M., Sherman, M. A., D’Gama, A. M., Doan, R. N., Rento, L. M., Girsakis, K. M., Bohrsen, C. L., Kim, S. N., *et al.* (2021) The landscape of somatic mutation in cerebral cortex of autistic and neurotypical individuals revealed by ultra-deep whole-genome sequencing. *Nat. Neurosci.*, 24, 176–185
  202. Fachal, L., Aschard, H., Beesley, J., Barnes, D. R., Allen, J., Kar, S., Pooley, K. A., Dennis, J., Michailidou, K., Turman, C., *et al.* (2020) Fine-mapping of 150 breast cancer risk regions identifies 191 likely target genes. *Nat. Genet.*, 52, 56–73
  203. Dzida, T., Iqbal, M., Charapitsa, I., Reid, G., Stunnenberg, H., Matarese, F., Grote, K., Honkela, A. and Rattray, M. (2017) Predicting stimulation-dependent enhancer-promoter interactions from CHIP-Seq time course data. *PeerJ*, 5, e3742
  204. Feng, Z. X., Li, Q. Z. and Meng, J. J. (2018) Recognition of the long range enhancer-promoter interactions by further adding DNA structure properties and transcription factor binding motifs in human cell lines. *J. Theor. Biol.*, 445, 136–150
  205. Hait, T. A., Elkon, R. and Shamir, R. (2022) CT-FOCS: a novel method for inferring cell type-specific enhancer-promoter maps. *Nucleic Acids Res.*, 50, e55