

journal homepage: www.elsevier.com/locate/csbj

Review

Exploring 3D chromatin contacts in gene regulation: The evolution of approaches for the identification of functional enhancer-promoter interaction

Hang Xu^a, Shijie Zhang^{a,b}, Xianfu Yi^c, Dariusz Plewczynski^{d,e}, Mulin Jun Li^{a,b,*}^a 2011 Collaborative Innovation Center of Tianjin for Medical Epigenetics, Tianjin Key Laboratory of Medical Epigenetics, Tianjin Medical University, Tianjin, China^b Department of Pharmacology, Tianjin Key Laboratory of Inflammation Biology, School of Basic Medical Sciences, Tianjin Medical University, Tianjin, China^c School of Biomedical Engineering, Tianjin Medical University, Tianjin, China^d Centre of New Technologies, University of Warsaw, Banacha 2c, 02-097 Warsaw, Poland^e Faculty of Mathematics and Information Science, Warsaw University of Technology, Koszykowa 75, 00-662 Warsaw, Poland

ARTICLE INFO

Article history:

Received 31 August 2019

Received in revised form 21 February 2020

Accepted 22 February 2020

Available online 28 February 2020

Keywords:

Chromatin Conformation Capture

Enhancer-promoter interaction

cis-Regulatory element

Chromatin loop

Machine learning

Computational method

ABSTRACT

Mechanisms underlying gene regulation are key to understand how multicellular organisms with various cell types develop from the same genetic blueprint. Dynamic interactions between enhancers and genes are revealed to play central roles in controlling gene transcription, but the determinants to link functional enhancer-promoter pairs remain elusive. A major challenge is the lack of reliable approach to detect and verify functional enhancer-promoter interactions (EPIs). In this review, we summarized the current methods for detecting EPIs and described how developing techniques facilitate the identification of EPI through assessing the merits and drawbacks of these methods. We also reviewed recent state-of-art EPI prediction methods in terms of their rationale, data usage and characterization. Furthermore, we briefly discussed the evolved strategies for validating functional EPIs.

© 2020 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Contents

1. Seeking functional EPIs	559
2. Identifying active CREs	560
3. Tracking spatial proximity	561
3.1. Crosslinking and proximity ligation introduce noises and artificial interactions	561
3.2. Restriction enzyme-based fragmentation limits the resolution of chromosome folding mapping	561
3.3. Loop calling methods are heterogeneous for significant interactions identification	562
3.4. Conventional 3C-based processing lacks ability to capture simultaneous or cooperative interactions	562
3.5. Current techniques face difficulties to exploit fine-scale interactions at single cells	562
4. Predicting unseen EPIs	562
4.1. Unsupervised learning methods	562
4.1.1. Distance-based methods	563
4.1.2. Correlation-based methods	564
4.1.3. Decomposition-based methods	565
4.2. Supervised learning methods	565
4.2.1. Training classifier with machine learning	565
4.2.2. Regression-based methods	566
4.3. Existing issues and challenges for EPI prediction	566
5. Validating functional EPIs and future direction	566

* Corresponding author at: Department of Pharmacology, Tianjin Key Laboratory of Inflammation Biology, School of Basic Medical Sciences, Tianjin Medical University, Tianjin, China.

E-mail address: mulinli@connect.hku.hk (M.J. Li).

Credit authorship contribution statement	567
Declaration of Competing Interest	567
Acknowledgments	567
Appendix A. Supplementary data	567
References	567

1. Seeking functional EPIs

In the last several decades, researchers have identified several types of functional DNA elements that can regulate tissue/cell type-specific gene expression in *cis* [1]. These *cis*-regulatory elements (CREs) are often located in the non-coding genomic regions, which make up of over 98% of the human genome [2]. Promoters and enhancers are two major CREs that control context-dependent gene transcription, with which promoters drive gene expression adjacent to transcription start sites (TSS) and enhancers faithfully orchestrate transcription from distal position regardless of orientation [3,4]. However, when and how these elements precisely regulate target gene expression are largely unknown.

Eukaryotic enhancers are bound by various transcription factors (TFs) when activated and upregulate target gene expression by forming chromatin loops with promoters [4,5]. The chromatin loops are mostly driven by persistent or transient cohesin extrusion and other unknown mechanisms [6,7]. The complexity of EPI formation and its functional implementation limits accurate detection of functional EPIs in high throughput manner. First, it is estimated that over 1,000,000 enhancers in the human genome, whereas the number of promoters, even if the alternative promoters are considered, is in the same order of magnitude as the number of gene transcripts [4,8–10]. Such great redundancy enables target genes to be regulated by different enhancers and ensures robust gene control at different conditions, but in turn complicates the detection of tissue/cell type-specific EPIs [11,12]. Second, enhancer is suggested to regulate gene expression by forming loops with promoters of target genes [13]. Though the distance among most EPIs is less than 200 kb, in extreme cases, enhancer

can locate over 1 Mb away from its regulatory targets [14]. It is estimated that only 40% enhancers can regulate their nearest genes, and others cannot be detected through the nearest-gene rule [11]. Third, interaction between enhancer and promoter does not necessarily mean functional causation. EPIs detected by close proximity usually represent more likely association than causation because of contradictory evidence showing the influence of EPIs on gene expression [15–17]. Fourth, the dynamics of EPI and corresponding maintenance of gene expressions could be partially explained by loop extrusion model [18,19], and how the functional EPIs are established and maintained remain to be fully addressed.

Currently, techniques based on chromosome conformation capture (3C-based techniques) are commonly used to identify EPIs in different throughputs [20]. However, the identification of true functional EPIs during development and homeostasis usually requires extra efforts, such as profiling chromatin states, tracking TF binding and quantifying eventual gene expression [21,22]. Simplistically, evidence from three aspects are suggested to define a functional EPI in common practice: 1) active or meaningful chromatin states; 2) close spatial proximity (although several lines of evidence indicate enhancers could control target gene expression independent of EPI [17,23]); and 3) positive transcription outcome (Fig. 1). By focusing on assessing the drawbacks and merits, in this review we summarize and categorize the technologies for identifying active CREs, detecting chromatin proximity and validating potential EPIs. We briefly introduce the roles of the evolved 3D genomic profiling assays for characterizing functional EPIs. Furthermore, we illustrate how the state-of-the-art computational methods were developed based on these functional genomics data.

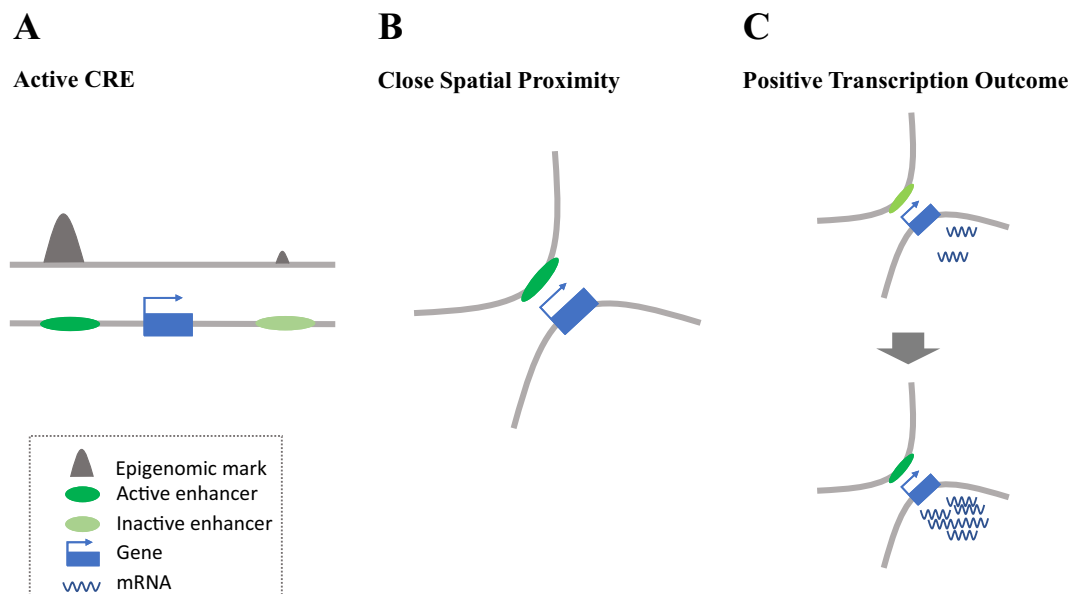


Fig. 1. Definition of functional EPIs. Functional EPIs required evidence from three aspects: (A) Active status of enhancers and promoters. (B) Spatial proximity between enhancer and promoter (though some studies revealed exceptional cases). (C) Context-dependent gene expression alteration.

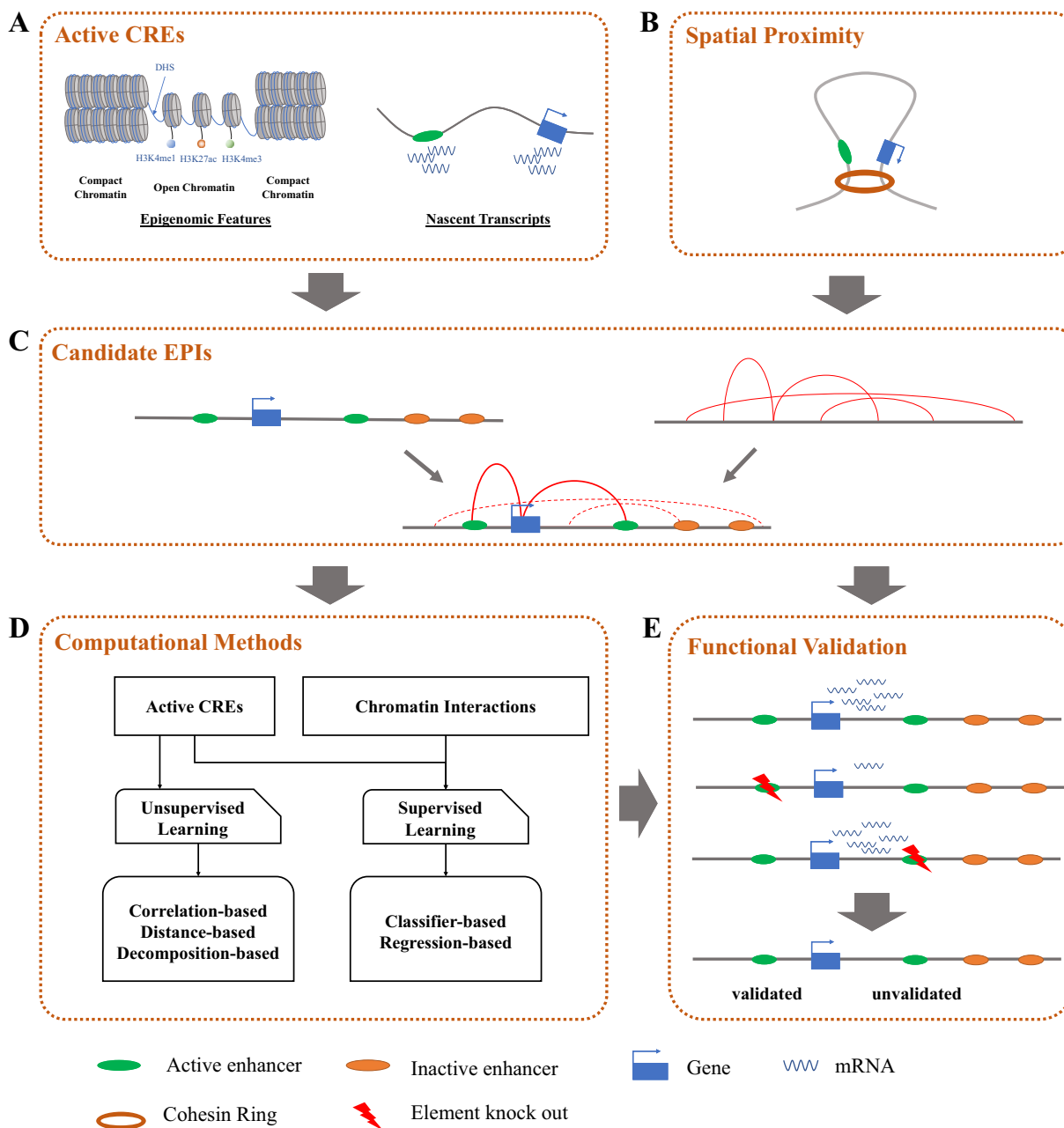


Fig. 2. Conventional workflow for detecting, predicting and validating functional EPIs. (A) Epigenomic features and nascent transcripts are the major characteristics of active CREs. (B) Functional EPIs require enhancer and promoter to be spatially adjacent. (C) Candidate EPIs are routinely derived from the combination of active CREs and chromatin loops. (D) Computational methods are developed on candidate EPIs using either supervised or unsupervised algorithms. (E) Disrupting CREs and testing the transcriptional effects on gene transcription are the main approaches to validate candidate EPIs.

2. Identifying active CREs

A key component of functional EPI detection is to investigate whether the associated CREs are activated in particular conditions. Comparative genome analysis has been widely applied in identifying functional DNA elements through searching the cross-species conserved regions [24,25]. But its application on enhancer detection is challenged since the findings that enhancers evolve rapidly and most of them are species-specific [12]. The lack of ability in measuring activity state of CREs is another limitation of comparative genome analysis, which promotes the development of new strategies to track active CREs beyond DNA sequence alone. Advances in biotechnologies and high-throughput sequencing greatly facilitate the identification of tissue/cell type-specific

enhancers and promoters. For examples, the binding of certain TFs, co-factors and histone modifiers (such as EP300, CDK7, BRD4, and Mediator) usually indicate active CREs. The histone modifications can mark the activity states of CREs (such as H3K27ac for active enhancer, H3K4me1 for primed enhancer, both H3K4me1 and H3K27me3 for poised enhancer, and H3K4me3 for active promoter [26–28]). These biological processes could be measured by ChIP-ChIP, ChIP-seq, as well as recent Cut&Run and Cut&Tag [29–32]. Active enhancers and promoters are also closely related to chromatin states (such as open chromatin and nucleosome occupancy), which is highly relevant to the binding of various TFs. Therefore, genomic assays measuring chromatin states, including DNase-seq, ATAC-seq and MNase-seq, are all informative for detecting active CREs [33–35]. But there is no one-to-one

correspondence between epigenomic feature and certain CRE. As a part of ENCODE project, ChromHMM and Segway are used to integrate multiple histone modifications and chromatin states across large numbers of tissues/cell types to generate comprehensive predictions for different CREs using DNA segmentation algorithms [36–38]. Detecting nascent RNAs is another feasible approach for the identification of active CREs. It is enlightened by the similar properties between enhancers and promoters. For instances, enhancers can be transcribed into non-coding enhancer RNAs (eRNA) when activated [39,40]. And promoters of some genes were found to have enhancer's ability and distally regulate the transcription of other genes [34,41,42]. Cap analysis of gene expression (CAGE) and similar techniques (like GRO-seq and PRO-seq) are very suitable for detecting transcripts within cell nucleus, thus are widely applied in detecting eRNA and nascent mRNA [43–47].

Studies have shown that poised enhancers can physically contact their target genes by polycomb dependent manner in certain cellular and genomic contexts, but the permissive EPIs will be silent until receiving active nuclear signatures [48,49]. Therefore, detecting active enhancers and promoters could be the prerequisite to define functional EPIs (Fig. 2A). Nevertheless, current understanding of CRE activation by specific chromatin marks or nascent transcripts do not necessarily imply that the enhancers or promoters are truly functional. For example, it was revealed that only 26% enhancers predicted in ENCODE project have regulatory activity [50]. Besides, recent CRISPR screening had uncovered that some regulatory regions with unmarked regulatory signatures are functional [51], which highlights the importance to exploit the novel chromatin features for active CRE definition. Taken together, these techniques do not provide direct evidence for the linkage between enhancers and promoters. Some patterns of the data, such as co-activation between enhancer and promoter across cells, are informative for functional EPI detection, which will be discussed further in the part of computational methods.

3. Tracking spatial proximity

Reduced spatial distance by chromatin loop formation is another critical property of functional EPI. 3C-based techniques lead to a revolution for identifying DNA interactions [52] (Fig. 2B). Through introducing proximity ligation to next generation sequencing (NGS), 3C and its derivatives are able to capture the three-dimensional interactions among chromatin. Especially, high-throughput 3C-derived techniques, including high-depth Hi-C, HiChIP and ChIA-PET using a large number of cells, provide efficient avenue to identify genome-wide potential EPIs to date [53–55]. Relying on modeling contact frequencies and assuming interaction background generated by random collisions across the chromatin polymer, significant interactions can be called through various computational methods [56]. However, conventional 3C-based techniques still have limitations in terms of resolution, sensitivity and expenditure. In fact, unless sustaining enough library complexity and investing ultra-high sequencing depth, the resolution of Hi-C was not precise enough to distinguish chromatin loops. A fundamental difficulty of 3C-based techniques was their dependence on proximity ligation. It was suggested that proximity ligation fails to capture many known structures and introduces high background noise [57], which significantly affects the quality of identified loops for follow-up EPI modeling. Besides, restriction enzyme dependent techniques could not distinguish genomic regions smaller than a theoretical limit determined by enzyme. For example, the theoretical maximum resolution of 6 bp restriction enzymes was 4 kb. ChIA-PET was suggested to have higher resolution with the same sequencing depth because it focused on regions marked by specific factors. But it was criticized to have

low sensitivity which leads to high false negative rate in detecting chromatin loops [55]. Additionally, it usually required a mass of cells to achieve high resolution for loop calling, which not only made it unaffordable for many studies, but also made it unable to detect chromatin interactions at single cell level. The development of single-cell or single-molecule 3C-techniques provide insights into single-cell loop calling, but most of them are immature and need to be validated and improved with more efforts [57]. Therefore, genome-wide identification of unbiased loops (including EPIs) is impossible by conventional 3C-based techniques. Some considerations and evolved strategies are briefly summarized here from different angles with previous reviews [58–61].

3.1. Crosslinking and proximity ligation introduce noises and artificial interactions

As two necessary steps in 3C library construction, crosslinking could capture unwanted contacts that are not mediated by direct chromatin interactions and introduce artifacts through intervening molecules or organelles, while ligation heavily relies on the specificity of crosslinked sequences and physicochemical state of chromatin in nucleus or solution. Although several modified techniques, such as DLO Hi-C [62] and BL-Hi-C [63], had been developed to optimize the efficiency and specificity during the crosslinking and ligation, the original defects still exist. Recently, several new methods complementary to previous 3C-based techniques have been invented to provide relatively unbiased investigation of chromosome interactions. Briefly, native 3C-based assay (i3C/iHi-C) captures spatial interactions without crosslinking [64,65]; genome architecture mapping (GAM) leverages ultrathin nuclear cryosectioning followed by sequencing to detect long distance and three-way contacts [66]; split-pool recognition of interactions by tag extension (SPRITE) discriminates different types of contacts by split-pool barcoding of DNA molecules within the same crosslinking complex [57]; transposase-mediated analysis of chromatin looping (Trac-looping) simultaneously detects chromatin accessibility and multiscale chromatin interactions without prior chromatin fragmentation and proximity-based ligation [67]; DNA adenine methyltransferase identification (DamID) named DamC provides the first crosslinking- and ligation-free demonstration of chromosome structure by DNA methylation-based detection [68]; multiplex chromatin-interaction analysis via droplet-based and barcode-linked sequencing (ChIA-Drop) captures complex chromatin interactions with single-molecule precision [60,69]. These novel methods that overcome some restrictions of conventional 3C-based library construction can achieve more reliable and powerful chromatin conformation measurement.

3.2. Restriction enzyme-based fragmentation limits the resolution of chromosome folding mapping

Theoretically, HaeIII and MboI can cut the human genome every 342 bp and 401 bp on average respectively [63]. Although increasing sequencing read coverage has been widely applied in Hi-C experiments, reliance on restriction enzyme limits the resolution of data despite read coverage. Factor-specific 3C-based techniques, like ChIA-PET, HiChIP and PLAC-seq on architecture proteins (such as CTCF and YY1), RNA Polymerases as well as histone modifications (such as H3K27ac and H3K27me3), improve the yield of conformation-informative reads and increase resolution by chromatin immunoprecipitation and peak calling [55,70–74]. In addition, Capture-C, Tiled-C, T2C, HiCap and other types of Capture Hi-C (CHI-C) enrich selected regions of interest by pre-designed capture oligonucleotides, which significantly increase power of interaction detection at high resolution [75–80]. Syn-HiC redesigns chromosome with regularly spaced restriction sites thus enables

unbiased distribution of contact frequencies and robust definition of Hi-C resolution [81]. Avoiding to use restriction enzyme, Micro-C and DNase Hi-C utilize micrococcal nuclease and DNase I respectively to achieve mononucleosome resolution [82,83].

3.3. Loop calling methods are heterogeneous for significant interactions identification

Many computational tools have been developed to call loops from genome-wide 3C-based data (notably Hi-C), and the performance of several representative methods was comprehensively benchmarked before [56,84]. Extreme difference in the number of identified interactions, varied mean distance between the interacting points, low reproducibility among replicates as well as high false discovery rate upon simulated data were observed for existing loop calling methods. Such defects could be attributed to several intrinsic issues of data. First, there is no internal normalization criteria or reference of ground truth to convert relative contact probabilities into absolutely comparable values, which complicates the assumption of background model for contact frequency and hampers the definition of significant threshold referring to true EPIs. Multiplexed single molecule FISH combined with recent optical reconstruction of chromatin architecture (ORCA) that identify kilobase-level EPIs could provide a potentially complementary reference dataset [85]. Second, 3C-based library construction causes insufficient power of loop calling methods to detect long range *cis* (>1 Mb on the same chromosome) and *trans* interactions [86]. Some targeted enrichment approaches, such as Chi-C, HiChIP and ChIA-PET, could preferentially capture long range interactions. Specifically, ligation-free strategies, like GAM, SPRITE, Trac-looping and DamC provide more adequate power to identify long range *cis* and *trans* loops [60].

3.4. Conventional 3C-based processing lacks ability to capture simultaneous or cooperative interactions

The redundancy of enhancers compared to promoters implies the relationship between enhancer and promoter cannot be 1-to-1, and the percentage of enhancers interacted with multiple promoters were estimated from 9% to 50% [14,87]. It was speculated that simultaneous interactions between enhancers and promoters are important to ensure stable gene expression in transcription factories [14,88]. Although conventional 3C-based techniques can be used to detect multiple-contacts, it significantly lacks throughput and resolution to locate CREs genome-wide [78,89]. To capture simultaneous interactions at scale, several multi-way chromatin contacts identification methods have been developed. For examples, tethered multiple 3C (TM3C), that maps genome-wide simultaneous chromatin contacts via ligation of fragments upon agarose gel beads followed by paired-end sequencing [90]. Chromosomal walks (C-walks) investigates higher-order organization by linking multiple genomic loci together into proximity chains [91]. GAM enables the detection of three-way chromatin contacts but the resolution (>100 kb) is not enough to capture the interactions between multiple enhancers and promoters [66]. Multi-contact 4C (MC-4C) applies nanopore sequencing to measure multi-way DNA conformations in individual alleles using modified 4C-seq method [92,93]. Similar to MC-4C, Tri-C efficiently detects multiple ligation junctions within single sonicated 3C fragments by oligonucleotide capture [87]. SPRITE can identify multiple loci that simultaneously interact in a single cluster and long-distance [57]. Likewise, Trac-looping provides unbiased detection of multiple-way chromatin interactions and captures chromatin interactions across extremely large distances [67]. Taken together, the development of multi-way methods with diverse strategies have made it possible to detect multi-contacts among enhancers and promoters.

3.5. Current techniques face difficulties to exploit fine-scale interactions at single cells

Given the rapid evolution of 3C-based techniques at multiple contacts level, most of them generally capture snapshots of 3D genome for the whole cell populations at specific time point. To describe the highly dynamic 3D chromatin at single cells, several groups have optimized the 3C-based techniques and applied them to track chromatin conformation on single cells across different development stages and conditions [94–98]. However, the genomic resolution of current single cell chromatin conformation methods is limited due to either the highly variable chromatin structure among cell populations or the technical issue of subsampling, which significantly prevent the identification of stable EPIs at single cell level. Despite all these, Dip-C improves the detection power of chromatin contacts by combining a transposon-based whole-genome amplification [99], while ChIA-Drop uses droplet-based chromatin interaction analysis and reveals many promoter-centered multivalent interactions at high resolution [69]. Such technical progresses will initiate the in-depth exploitation of fine-scale EPIs at single cell level.

4. Predicting unseen EPIs

The continuous evolution of proximity-based 3D genome techniques made them possible to detect EPIs in genome wide. Meanwhile, accumulated genomic, epigenomic, and transcriptomic profiling data provided abundant resources for the identification of active CREs. To detect meaningful EPIs in particular conditions, the common practice would be to overlap high resolution chromatin interactions with tissue/cell type-specific active CREs (Fig. 2C) [20]. Although the discovery of EPIs has been fueled by integrating tissue/cell type-specific epigenomic profiles and 3D genomic data, high resolution genome-wide loop data are only available for limited human tissues/cell types thus far [100,101]. Unless the limitations such as low resolution, low sensitivity and high cost were properly addressed, 3C-based techniques could not be widely applied in most studies for chromosome loop identification. Besides, EPI was not the exclusive case in terms of spatial neighborhood, thus the chromatin loops identified by 3C data could be interpreted as either functional loops or other interactions by chance. To predict the unrecognized EPIs at different contexts, many computational methods have been developed by learning or modeling existing 3D genomic data and other molecular phenotype profiles, such as open chromatin, transcript expression, histone modification and TF binding [102,103] (Fig. 2D). Pioneered by epigenetic mark-promoter linkage studies [28,104], over 30 *in silico* methods currently have been proposed to predict EPIs in human using diverse omics datasets and statistical models. Generally, existing computational EPI prediction methods could be divided into two major categories including unsupervised and supervised learning (Table 1).

4.1. Unsupervised learning methods

Attributing enhancers to their nearest genes was commonly used approach to identify EPIs. Besides, increasing genomic/epigenomic features provided another approach to detect EPIs since active enhancers and promoters that have distinct patterns compared to inactive ones. These characteristics further enable the correlation of epigenomic signals between enhancer and promoter to be useful criterion for identifying potential interactions [28]. Similarly, the relation between gene expression and EPIs was also a feasible strategy to detect active EPIs, which makes use of linear/non-linear regression to quantitatively estimate the regulatory

Table 1
Computational methods for EPI prediction.

Tool	Year	Method category	Features	Algorithm	Links
Ernst et al. [28]	2011	Correlation-based	Histone marks, TF binding	Pearson's Correlation	http://compbio.mit.edu/ENCODE_chromatin_states/
Thurman et al. [104]	2012	Correlation-based	DHS	Pearson's Correlation	https://genome.ucsc.edu/ENCODE/downloads.html
DRE-target [115]	2013	Correlation-based	DHS, Sequence homology	Pearson's Correlation	ftp://public:public@202.120.224.143/NAR2013.tar.gz
Andersson et al. [11]	2014	Correlation-based	CAGE	Pearson's Correlation	http://fantom.gsc.riken.jp/5/
PreSTIGE [106]	2014	Distance-based	Distance, Insulator	Linear Domain Models	http://mendel.gene.cwru.edu:8080/
gkm-SVM [139]	2014	Train Classifier	DNA	Support Vector Machine	http://www.beerlab.org/gkmsvm/
IM-PET [123]	2014	Train Classifier	Histone marks, TF binding, DNA, RNA-seq	Random Forest	www.healthcare.uiowa.edu/labs/tan/IM-PET_Package.tgz
ELMER [114]	2015	Correlation-based	DNA methylation, RNA-seq	Pearson's Correlation	https://static-content.springer.com/esm/art%3A10.1186%2Fs13059-015-0668-3/MediaObjects/13059_2015_668_MOESM4_ESM.xlsx
RIPPLE [125]	2015	Train Classifier	Histone marks, TF binding, DHS, DNA-seq	Random Forest	http://pages.discovery.wisc.edu/~sroy/ripple/index.html
PEGASUS [116]	2015	Correlation-based	Conservation	Linkage Scoring	ftp://ftp.biologie.ens.fr/pub/dyogen/PEGASUS/
Basset [140]	2016	Train Classifier	DNA	CNN	https://github.com/davek44/Basset
TargetFinder [126]	2016	Train Classifier	Histone marks, TF binding, DHS, CAGE	Gradient Tree Boosting	https://github.com/shwhalen/targetfinder
PETModule [124]	2016	Train Classifier	Histone marks, Conservation, Motif	Random Forest	http://hulab.ucf.edu/research/projects/PETModule/
EpiTensor [119]	2016	Decomposition-based	Histone marks	Tensor Decomposition	http://wanglab.ucsd.edu/star/EpiTensor/
JEME [141]	2017	Regression-based	Histone marks, DHS, DNA methylation, eRNA	Linear Regression	https://github.com/yiplabcuhk/JEME
McEnhancer [135]	2017	Train Classifier	DHS	Markov Chain Model	https://ohlerlab.mdc-berlin.de/software/McEnhancer_134/
SWIPE-NMF [120]	2017	Decomposition-based	eQTL, DHS	Matrix Factorization	https://github.com/kaiyuanmifeng/SWIPE-NMF
EPIANN [130]	2017	Train Classifier	DNA	CNN + Attention Model	https://github.com/wgmao/EPIANN
PEP [131]	2017	Train Classifier	DNA	Gradient Tree Boosting	https://github.com/ma-compbio/PEP
CISD [138]	2017	Train Classifier	MNase-seq	Logistic Regression	https://github.com/huizhangucas/CISD
FOCS [142]	2018	Regression-based	DHS, CAGE, GRO-seq	Linear Regression	https://github.com/Shamir-Lab/FOCS
Cicero [111]	2018	Correlation-based	scATAC-seq	Graphical Lasso	https://github.com/cole-trapnell-lab/cicero-release
TransDecomp [121]	2018	Decomposition-based	CAGE	Decomposition	https://github.com/anderssonlab/transcriptional_decomposition
Rambutan [136]	2018	Train Classifier	DNA, DHS	CNN	https://github.com/jmschrei/rambutan
SPEID [129]	2018	Train Classifier	DNA	CNN	https://github.com/ma-compbio/SPEID
3DEpiLoop [127]	2018	Train Classifier	Histone marks, TF binding	Random Forest	https://bitbucket.org/4dnucleome/3depiloop
EP2vec [132]	2018	Train Classifier	DNA	Word2vec + Gradient Boosted Regression Trees	https://github.com/wanwenzeng/ep2vec
DeepTACT [137]	2019	Train Classifier	DHS, DNA	CNN + Attention Model	https://github.com/liwenran/DeepTACT
C3D [112]	2019	Correlation-based	DHS	Pearson's Correlation	https://github.com/LupienLab/C3D
EPIP [128]	2019	Train Classifier	DHS, Histone marks	Adaboost	http://www.cs.ucf.edu/~xiaoman/EPIP/
DRAGON [152]	2019	Polymer Simulation	Histone marks, TF binding	Maximum Entropy	https://github.com/ZhangGroup-MITChemistry/DRAGON
CHINN [133]	2019	Train Classifier	DNA	CNN	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE135052
CT-FOCS [143]	2019	Regression-based	DHS	Linear Mixed Effect Models	http://acgt.cs.tau.ac.il/ct-focs
HiC-Reg [150]	2019	Regression-based	DHS, Histone marks, TF binding	Random Forests	https://github.com/Roy-lab/HiC-Reg
ABC [110]	2019	Distance-based	Distance, DHS, Histone marks	Activity-by-contact Model	https://github.com/broadinstitute/ABC-Enhancer-Gene-Prediction
3DPredictor [146]	2020	Train Classifier	CAGE, CTCF	Gradient Boosting	https://github.com/labdevgen/3DPredictor

potential of enhancers towards given genes. In general, these unsupervised methods could be classified into three categories: (1) distance-based methods; (2) correlation-based methods and (3) decomposition-based methods (Fig. 3A–C).

4.1.1. Distance-based methods

Linking enhancers to the nearest promoter had been widely used in many studies. Despite the simple rationale, methods applying this strategy was proved to be considerably effective. It was

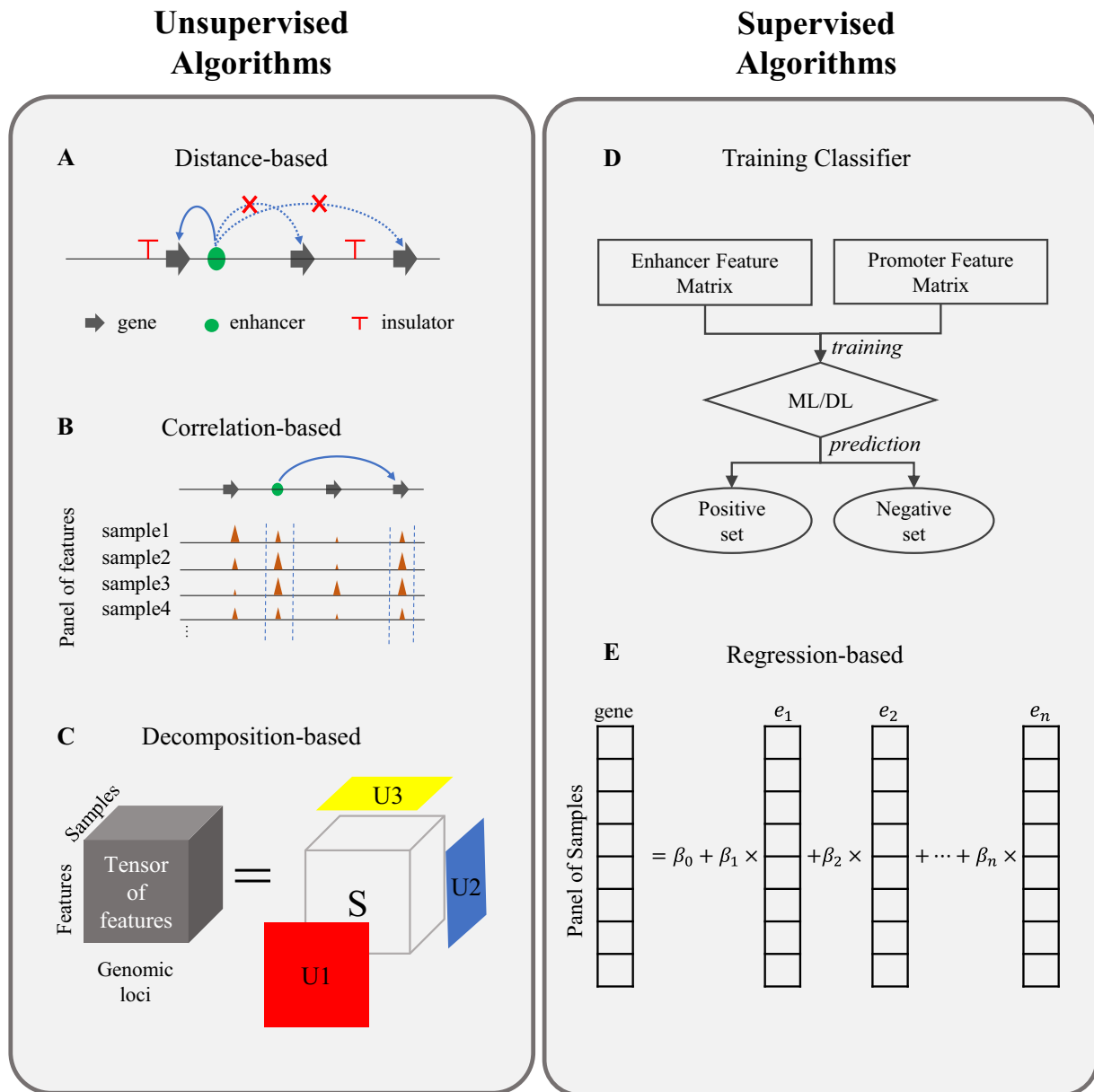


Fig. 3. Overview of computational methods for EPI prediction. Strategies to predict EPIs can be divided into two major categories, unsupervised learning and supervised learning. Unsupervised learning algorithms include (A) Distance-based methods assign enhancers to the nearest genes, and the regulatory scope is restricted in some methods. (B) Correlation-based methods detect EPI according to high correlation of chromatin features between enhancer and promoter from a panel of samples. (C) Decomposition-based methods decompose feature matrix/tensor into subspaces, which capture the spatial features of genome thus could be used to detect EPI. Supervised learning algorithms include (D) Training Classifier methods measure the relationship between gene activity and enhancers by estimating the regulatory potential of enhancers for specific gene. (E) Regression-based methods build different machine learning classifier to distinguish positive EPIs from randomly selected negative set. ML: machine learning, DL: deep learning.

estimated that in 40% cases enhancers regulate the nearest genes [11,105]. Setting additional criteria can increase the efficiency of distance-based methods. For example, considering the cell type-specific activated enhancer and promoters, PreSTIGE captured EPIs that only exist in specific cell types [106]. In addition to distance, false positive rate caused by distance-based methods were greatly reduced by constraining the range of EPIs within topologically associating domains (TADs) [107] or insulated neighborhoods (INs) [108] which limit the regulatory potential of enhancers within specific 3D genome architecture [109]. Recent activity-by-contact (ABC) model scored the potential of EPI by combining distance effect and enhancer activity, demonstrating a superior performance than existing approaches [110]. Nevertheless, as a naïve strategy that barely consider long distance interaction,

distance-based methods were commonly used as baseline approach to predict EPIs in many studies and the performance of distance-based methods was usually not as good as subsequent methods.

4.1.2. Correlation-based methods

Many EPI prediction methods, including Ernst et al. [28], Thurman et al. [104], Cicero [111] and C3D [112] were developed based on the correlation of epigenomic marks between enhancers and promoters. To be specific, Ernst et al. calculated Pearson's correlation of histone modifications and TF binding between enhancers and approximate promoters [28], while, Thurman et al. leveraged DNase I hypersensitive sites (DHSs) across many cell types to compute correlation [104]. Similarly, C3D was able to detect correlated

CREs with more DHS data [112]. Besides, with single cell ATAC-seq [35] profiles, Cicero was designed to detect correlation of open chromatin at single cell level [111]. Instead, Andersson et al. [11] used cap analysis gene expression (CAGE) [113] profiles to inspect whether consistent transcriptional events could be observed between enhancer locus (such as enhancer RNA (eRNA)) and target genes across different tissues/cell types. ELMER uses inverse correlation between DNA methylation and expression of nearby genes to predict transcriptional targets [114]. DRE-target [115] identified target genes of distal regulatory elements (DREs) as those obtaining high phylogenetic correlation with DREs. Despite the improved performance, DRE-target depended on Hi-C data which were not widely available. Likewise, PEGASUS relied on evolutionary conservation of synteny to estimate enhancer-gene associations [116,117]. The performance of correlation-based methods was not only influenced by the choice of features, it was also suggested to be affected by the algorithm to calculate correlation [111]. Pearson's correlation was very sensitive to outliers, usually generating large numbers of false positive predictions. Among the correlation-based EPI methods, only Cicero addressed this problem by using graphical Lasso [111] which calculated regularized correlation matrices, thus was more robust to outliers [118].

4.1.3. Decomposition-based methods

With accumulating genomic/epigenomic/transcriptomic data from diverse tissues/cell types, matrix decomposition became a feasible approach to detect EPIs by extracting meaningful co-variation patterns from high-dimensional signals. Hitherto there were three EPI methods based on matrix decomposition, including EpiTensor [119], SWIPE-NMF [120] and TransDecomp [121]. EpiTensor collected various assays from many cell types and combined all the data with a three-order tensor in which the dimensions represent genomic loci, assay type and cell type respectively. Tensor decomposition was then applied to resolve the combined tensor into three subspaces: cell subspace, assay subspace and locus subspace. By analyzing the eigenvectors of locus subspace, genomic interactions were captured by linking the peaks of eigenvectors following distance-based approaches [119]. SWIPE-NMF firstly included six types of genomic segments and then established association matrices for every pairs of segments. Extended version of three-factor penalized matrix factorization (PMF) was then used to factorize the association matrices. Enhancer-promoter interactions were then characterized from the results of PMF [120]. TransDecomp implemented very different strategies compared with EpiTensor and SWIPE-NMF. It collected CAGE signals and decomposed the data to get two principle components called positional independent (PI) and positional dependent (PD). Then TransDecomp set 25 features related to the components to train a random forest classifier to identify EPIs derived from promoter-capture Hi-C [122]. The results showed that features derived from PI and PD components were unique in distinguishing active EPIs [121]. Different with correlation-based methods that only rely on limited data type across large numbers of samples, the decomposition-based methods leveraged multi-scale information from omics data to learn unique patterns for putative EPIs.

4.2. Supervised learning methods

With the development of 3C-based techniques, especially the high throughput methods, such as high-depth Hi-C, HiChIP and ChIA-PET in large numbers of cells, EPIs could be effectively described across entire genome, which made it possible to use supervised methods to identify potential EPIs (Fig. 3D and E).

4.2.1. Training classifier with machine learning

There were many efforts to detect EPIs by training classifiers. By leveraging various 1D genomic/epigenomic features, classifiers could be established to distinguish Hi-C/ChIA-PET supported EPIs from random selected negatives. IM-PET was the first method following this strategy [123]. It implemented four features, including enhancer and promoter activity profile correlation, transcription factor and target promoter correlation, coevolution of enhancer and target promoter and distance constraint between enhancer and target promoter [123]. Therefore, IM-PET was a combination of several strategies mentioned above, but this made it not user-friendly since the way of calculating those integrated features was not easy to fix, which raises additional challenges. PETModule implemented a similar feature sets with IM-PET but showed higher performance [124]. Instead, in subsequent studies, such as RIPPLE [125], TargetFinder [126], 3DEpiLoop [127] and EPIP [128] directly used epigenomic/transcriptomic profiles, including TF binding, histone marks, DHSs and expression data to train comprehensive classifier. Both RIPPLE and TargetFinder carried out careful feature evaluation and identified expression level as the most distinct feature, and DHS, CTCF binding were also informative in trained models. Besides, TargetFinder included ChIP-seq data for over 100 TFs and the feature importance of those data was top-ranked. Both RIPPLE and TargetFinder achieved high performance in identifying EPIs within specific cell types. However, the requirement of plenty of features made it impractical to apply RIPPLE and TargetFinder to train specific classifiers for more tissues/cell types.

Efforts had also been put to explore the possibility of EPIs detection with sequence features only. There were several methods developed based on this approach, including SPEID [129], EPIANN [130], PEP [131], EP2vec [132] and CHINN [133]. DNA sequences were usually represented with one-hot encoding, but with the development of deep learning methods, especially the great success of word2vec in nature language processing, word embedding, like dna2vec, had been considered in motif feature analysis [129–132]. Among those methods, SPEID and EPIANN used one-hot encoding while PEP and EP2vec tried dna2vec approach to represent DNA sequence. All sequence-based methods focused on efficiently extracting the information from DNA sequences. To pinpoint it, SPEID implemented a bidirectional long short-term memory (BiLSTM) [134] module before training classifier, while EPIANN applied attention mechanism to directly locate the functional DNA elements. PEP and EP2vec trained a word embedding model first, and then trained classifiers with embedded DNA sequences. In addition, to identify cell type-specific EPIs, sequence-based methods required active enhancers and promoters to be well defined in given condition, such as CHINN used DNA sequences of the interacting open chromatin regions [133].

To overcome the limitation of sequence-based methods, some algorithms used widely available open chromatin profiles to supply cell type-specific signatures. For example, McEnhancer [135] learned related DHS-gene pairs from small number of known pairs and used a semi-supervised strategy to predict unlabeled ones. Besides, Rambutan [136] used convolutional neural network (CNN) to extract features from both DNA and DNase-seq data, and the summarized features together with the distance between enhancer and promoter were finally used to make predictions. Similarly, DeepTACT [137] applied CNN layers to extract features from raw data, but the merged features were further processed with BiLSTM and attention layers to make better integration. Other type of chromatin accessibility data, such as MNase-seq could also be used in EPI prediction [34]. CISD implemented MNase-seq to train classifier to distinguish EPIs derived from ChIA-PET loops against randomly selected ones [138].

According to different algorithms aforementioned, statistical learning methods could be divided into two categories: typical

machine learning and deep learning. Machine learning methods used random forest, logistic regression, and support vector machine (SVM) to establish classifiers directly with Hi-C/ChIA-PET loops and various feature sets. According to RIPPLE, ensemble learning methods, such as random forest, usually performed better than other classification methods [125]. However, feature selection could not be performed by random forest. To solve this problem, RIPPLE combined random forest with Lasso for feature selection. On the other hand, deep learning methods used CNN, recurrent neural network (RNN) and attention model to extract informative features from raw input and then implemented simple machine learning to train classifiers. Deep learning methods usually performed better when sample size is large enough. Specifically, when implementing attention layers, learning kernel could locate the relatively precise position of enhancer elements [23,130].

Interestingly, some machine learning methods which were not originally designed for EPI prediction could also be applied to this topic. For example, gkm-SVM [139] and Basset [140] were developed to identify regulatory elements, but when the input was set as the EPI samples and concatenated features of enhancers and promoters, both gkm-SVM and Basset were easily transformed into EPI prediction methods.

4.2.2. Regression-based methods

Although training classifiers to distinguish active EPIs from others was a promising strategy, the regulatory potential of specific enhancers to their target genes could not be properly quantified. This limitation could be addressed by regression-based methods. When considering gene expression level as the effect indicator of neighboring CREs, the regulatory ability of separate enhancers could be inferred from training regression model among a panel of expression data together with epigenomic profiles in many cell types. In other words, the parameters learned from regression model may represent the degree of enhancers influencing target genes. Based on this rationale, JEME [141] quantitatively estimated the universal regulatory potential of enhancers while cell type-specific EPIs were then identified by training a random forest classifiers. Similar to JEME, FOCS [142] implemented a leave-n-out algorithm to obtain robust regression model linking nascent RNA transcription (measured by GRO-seq [47]) with DHSs data for correlated enhancer and promoter activity across many samples. CT-FOCS extended the FOCS model to use multiple replicates per cell type to infer cell type-specific EPIs [143]. Therefore, compared with the classifier learning, the regression-based methods can properly evaluate regulatory potential by directly associating CREs with gene transcription.

4.3. Existing issues and challenges for EPI prediction

Despite the great achievements, there are limitations among strategies applied in the current state-of-art computation methods. Distance-based methods highly depend on accurate identification of cell type-specific regulatory elements, which usually requires multiple genomic and epigenomic features. Distance-based methods are also hampered by the fact that many enhancers regulate distal promoters. Correlation-based methods and regression-based methods have higher performance, but they require large panel of samples, thus is usually not generalizable when unseen cell type is given. In contrast, supervised learning methods which train classifiers are applicable in predicting EPI across cell types. The merits and disadvantages of each EPI prediction strategy have been discussed before [102,103]. Here, we supplemented some critical problems that significantly affect the performance of current methods and briefly summarized related challenges of future prediction task.

First, there is no widely accepted ground truth to systematically evaluate the existing methods, probably due to the low power and high false positive rate of current technologies in capturing chromatin interactions. A recent benchmark study generated a set of candidate enhancer-gene interactions (BENGI) by integrating the candidate active CREs with experimentally validated genomic interactions on specific tissues/cell types [144]. They found that, overall, the correlation-based methods did not outperform the distance-based methods, and the supervised methods, like TargetFinder [126], were only modestly better than distance-based methods for most benchmark datasets when trained and tested with the same cell type but underperformed when applied across cell types. Second, the inflated performance of supervised learning methods was frequently observed [145]. The key problem could be attributed to the improper generation of training dataset and biased sampling procedure, in which the positive samples shared same features but negative samples obtained varied feature distribution. Random-split strategy in preparation of training dataset seems to overcome such problems [145,146]. Third, methods using deep learning have problems in doing feature selection. To address this problem, SPEID evaluated feature importance by measuring the decreasing of method performance when replacing certain feature value with random noise [129]. To evaluate feature importance in deep learning-based methods, one can use the feature selection strategies including SHAP [147], DeepLIFT [148], and Deep Feature Selection [149]. Fourth, most of existing computational methods modeled the chromatin interaction by either spatial proximity or transcriptional outcome, which makes them face difficulty in verifying causal relationship instead of showing only correlation. Although methods, such as HiC-Reg [150], 3Dpredictor [146], MEGABASE [151], DRAGON [152] and ABC [110], can quantitatively measure the interacting probability or intensity, how such prediction values are proportional to functional readout is largely unknown. Recent biochemical experiments showed that the increased EPI could lead to decreased gene activation [17,19,23], which further complicates the establishment of causal link between spatial proximity and transcriptional outcome. Lastly, computational EPI prediction methods are facing challenges and new opportunity brought by advanced techniques. For example, some 3D genomic techniques, such as C-walks [91], GAM [66], MC-4C [92], SPRITE [57] and ChIA-drop [69], have been developed to detect multiplex chromatin interactions in single allele. In these data, multiple *cis*-regulatory elements can interact with same target gene simultaneously, which is difficult scenario for current EPI methods, especially for methods implemented training classifiers. Methods based on regression could detect multiple interactions theoretically [153,154], but whether the interactions are active simultaneously or just gloss caused by the average of bulk cells remain elusive.

5. Validating functional EPIs and future direction

Even if the advance of biochemical and computational methods has deepened our understanding of precise transcriptional control in the 3D genome, the biological function of CREs as well as the links between enhancer and their regulated targets identified by current techniques remain to be validated (Fig. 2E). Using transgenic reporter assays and massively parallel reporter assays, candidate enhancers could be dissected regardless of their native genomic context and endogenous target genes [5,155,156]. *In vivo* enhancer manipulation can be performed by Clustered Regulatory Interspaced Short Palindromic Repeat (CRISPR) associated Cas (CRISPR/Cas) system currently. Several lines of CRISPR/Cas strategies, including nuclease-active genome-editing screens and nuclease-inactive epigenome-editing screens, have been success-

fully applied to characterize large numbers of enhancers in their native genomic context [5]. On the other hand, super-resolution DNA FISH and microscopy complement with proximity-based techniques provide an unprecedented view of chromatin interactions at kilobase-scale resolution [157]. However, accurate identification of active CREs and bona fide EPIs in particular condition are usually two uncoupled experiments, which cannot answer which EPIs are true functional and eventually modulate transcriptional event of target genes or other molecular phenotypes. The objective of validating true functional EPIs have sparked enormous interest in designing novel coherent experiment. For examples, by introducing chromatin loops at desired genomic loci, chromatin loop reorganization using CRISPR-dCas9 (CLOuD9) can selectively inspect gene expression at targeted loci [158]; CRISPR affinity purification *in situ* of regulatory elements (CAPTURE) simultaneously identifies locus-specific transcriptional regulator complexes, chromatin-associated RNA and DNA interactions [159]; CRISPR-genome organization (CRISPR-GO) system can efficiently control the spatial positioning of genomic loci relative to specific nuclear compartments, enabling interrogation of chromatin interaction dynamics and associated molecular events [160]; light-activated-dynamic-looping (LADL) system allows light-inducible loop formation followed by single-molecule RNA-FISH for nascent expression quantification [161]. These new technologies greatly facilitate the one-stop evaluation of true biological functions for individual EPIs.

The evolution of advanced biotechnologies and accumulated functional genomics data are constantly revolutionizing the genome-wide identification of functional EPIs. By integrating multilayer tissue/cell type-specific evidence from uncoupled assays on genomic, epigenomic and transcriptomic profiling, the false positive rate of functional EPIs discovery could be reduced [162]. For example, tissue/cell type-specific quantitative trait locus mapping on gene expression (eQTL), chromatin accessibility (caQTL) and promoter interaction (pieQTL) have been used to refine or ascertain true functional EPIs together with active CREs profiling and 3C-based techniques [77,163–166], which will ultimately facilitate the interpretation of non-coding regulatory variant effect on 3D genome and complex disease [167]. In addition, the high throughput CRISPR/Cas-based perturbation screenings, such as Mosaic-seq [168], crisprQTL mapping [169] and CRISPRi-FlowFISH [110], on multiple target genes have offered promising strategies to simultaneously validate the endogenous effect of CREs with their putative target genes. Although these CRISPR/Cas-based systems are still in its infancy which only identified several hundreds of high-confidence EPIs, we believe they will initialize fundamentally novel computational methods by combining advanced 3D genomic data in the near future.

Credit authorship contribution statement

Hang Xu: Investigation, Writing - original draft. **Shijie Zhang:** Writing - review & editing. **Xianfu Yi:** Writing - review & editing. **Dariusz Plewczynski:** Writing - review & editing. **Mulin Jun Li:** Conceptualization, Investigation, Supervision, Writing - original draft.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by grants from the National Natural Science Foundation of China 31871327 (M.J.L.), Natural Science

Foundation of Tianjin 19JCJC63600 (M.J.L.); and Polish National Science Centre (2014/15/B/ST6/05082) and Foundation for Polish Science co-financed by the European Union under the European Regional Development Fund (D.P.).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.csbj.2020.02.013>.

References

- [1] Kellis M, Wold B, Snyder MP, Bernstein BE, Kundaje A, Marinov GK, et al. Defining functional DNA elements in the human genome. *Proc Natl Acad Sci U S A* 2014;111:6131–8.
- [2] Maston GA, Evans SK, Green MR. Transcriptional regulatory elements in the human genome. *Annu Rev Genomics Hum Genet* 2006;7:29–59.
- [3] Lenhard B, Sandelin A, Carninci P. Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nat Rev Genet* 2012;13:233–45.
- [4] Ong CT, Corces VG. Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nat Rev Genet* 2011;12:283–93.
- [5] Gasperini M, Tome JM, Shendure J. Towards a comprehensive catalogue of validated and target-linked human enhancers. *Nat Rev Genet* 2020.
- [6] Vian L, Pekowska A, Rao SSP, Kieffer-Kwon KR, Jung S, Baranello L, et al. The Energetics and Physiological Impact of Cohesin Extrusion. *Cell* 2018;173:1165–78 e20.
- [7] Davidson IF, Bauer B, Goetz D, Tang W, Wutz G, Peters JM. DNA loop extrusion by human cohesin. *Science* 2019;366:1338–45.
- [8] Landry JR, Mager DL, Wilhelm BT. Complex controls: the role of alternative promoters in mammalian genomes. *Trends Genet* 2003;19:640–8.
- [9] Davuluri RV, Suzuki Y, Sugano S, Plass C, Huang THM. The functional consequences of alternative promoter use in mammalian genomes. *Trends Genet* 2008;24:167–77.
- [10] Yoo EJ, Cooke NE, Liebhaber SA. Identification of a secondary promoter within the human B cell receptor component gene hCD79b. *J Biol Chem* 2013;288:18353–65.
- [11] Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, et al. An atlas of active enhancers across human cell types and tissues. *Nature* 2014;507:455–61.
- [12] Long HK, Prescott SL, Wysocka J. Ever-changing landscapes: transcriptional enhancers in development and evolution. *Cell* 2016;167:1170–87.
- [13] Bulger M, Groudine M. Functional and mechanistic diversity of distal transcription enhancers. *Cell* 2011;144:327–39.
- [14] Chepelev I, Wei G, Wangsa D, Tang Q, Zhao K. Characterization of genome-wide enhancer-promoter interactions reveals co-expression of interacting genes and modes of higher order chromatin organization. *Cell Res* 2012;22:490–503.
- [15] Osterwalder M, Barozzi I, Tissieres V, Fukuda-Yuzawa Y, Mannion BJ, Afzal SY, et al. Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature* 2018;554:239–43.
- [16] Symmons O, Pan L, Remeseiro S, Aktas T, Klein F, Huber W, et al. The Shh topological domain facilitates the action of remote enhancers by reducing the effects of genomic distances. *Dev Cell*. 2016;39:529–43.
- [17] Benabdallah NS, Williamson I, Illingworth RS, Kane L, Boyle S, Sengupta D, et al. Decreased enhancer-promoter proximity accompanying enhancer activation. *Mol Cell* 2019;76:473–84 e7.
- [18] Schwarzer W, Abdennur N, Goloborodko A, Pekowska A, Fudenberg G, Loe-Mie Y, et al. Two independent modes of chromatin organization revealed by cohesin removal. *Nature* 2017;551:51–6.
- [19] Rao SSP, Huang SC, Glenn St Hilaire B, Engreitz JM, Perez EM, Kieffer-Kwon KR, et al. Cohesin loss eliminates all loop domains. *Cell* 2017;171:305–20 e24.
- [20] Mora A, Sandve GK, Gabrielsen OS, Eskeland R. In the loop: promoter-enhancer interactions and bioinformatics. *Brief Bioinform* 2016;17:980–95.
- [21] Kim S, Shendure J. Mechanisms of Interplay between Transcription Factors and the 3D Genome. *Mol Cell*. 2019;76:306–19.
- [22] Stadholders R, Filion GJ, Graf T. Transcription factors and 3D genome conformation in cell-fate decisions. *Nature* 2019;569:345–54.
- [23] Alexander JM, Guan J, Li B, Maliskova L, Song M, Shen Y, et al. Live-cell imaging reveals enhancer-dependent Sox2 transcription in the absence of enhancer proximity. *eLife* 2019;8.
- [24] Mouse Genome Sequencing C, Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, et al. Initial sequencing and comparative analysis of the mouse genome. *Nature* 2002;420:520–62.
- [25] Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* 2005;15:1034–50.
- [26] Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* 2011;470:279–83.

- [27] Local A, Huang H, Albuquerque CP, Singh N, Lee AY, Wang W, et al. Identification of H3K4me1-associated proteins at mammalian enhancers. *Nat Genet* 2018;50:73–82.
- [28] Ernst J, Kheradpour P, Mikkelsen TS, Shores N, Ward LD, Epstein CB, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 2011;473:43–9.
- [29] Kim TH, Barrera LO, Zheng M, Qu C, Singer MA, Richmond TA, et al. A high-resolution map of active promoters in the human genome. *Nature* 2005;436:876–80.
- [30] Robertson G, Hirst M, Bainbridge M, Bilenky M, Zhao Y, Zeng T, et al. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat Methods* 2007;4:651–7.
- [31] Skene PJ, Henikoff S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife* 2017;6.
- [32] Kaya-Okur HS, Wu SJ, Codomo CA, Pledger ES, Bryson TD, Henikoff JG, et al. CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat Commun* 2019;10:1930.
- [33] Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, et al. High-resolution mapping and characterization of open chromatin across the genome. *Cell* 2008;132:311–22.
- [34] Kundaje A, Kyriazopoulou-Panagiotopoulou S, Libbrecht M, Smith CL, Raha D, Winters EE, et al. Ubiquitous heterogeneity and asymmetry of the chromatin environment at regulatory elements. *Genome Res* 2012;22:1735–47.
- [35] Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* 2013;10:1213–8.
- [36] Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* 2012;9:215–6.
- [37] Hoffman MM, Buske OJ, Wang J, Weng Z, Bilmes JA, Noble WS. Unsupervised pattern discovery in human chromatin structure through genomic segmentation. *Nat Methods* 2012;9:473–6.
- [38] Consortium EP. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012;489:57–74.
- [39] Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, et al. Widespread transcription at neuronal activity-regulated enhancers. *Nature* 2010;465:182–7.
- [40] Wang D, Garcia-Bassets I, Benner C, Li W, Su X, Zhou Y, et al. Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature* 2011;474:390–4.
- [41] Dao LTM, Galindo-Albarran AO, Castro-Mondragon JA, Andrieu-Soler C, Medina-Rivera A, Souaid C, et al. Genome-wide characterization of mammalian promoters with distal enhancer functions. *Nat Genet* 2017;49:1073–81.
- [42] Engreitz JM, Haines JE, Perez EM, Munson G, Chen J, Kane M, et al. Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature* 2016;539:452–5.
- [43] Shiraki T, Kondo S, Katayama S, Waki K, Kasukawa T, Kawaji H, et al. Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proc Natl Acad Sci U S A* 2003;100:15776–81.
- [44] Hirabayashi S, Bhagat S, Matsuki Y, Takegami Y, Uehata T, Kanemaru A, et al. NET-CAGE characterizes the dynamics and topology of human transcribed cis-regulatory elements. *Nat Genet* 2019;51:1369–79.
- [45] Churchman LS, Weissman JS. Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* 2011;469:368–73.
- [46] Mahat DB, Kwak H, Booth GT, Jonkers IH, Danko CG, Patel RK, et al. Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). *Nat Protoc* 2016;11:1455–76.
- [47] Core LJ, Waterfall JJ, Lis JT. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* 2008;322:1845–8.
- [48] Cruz-Molina S, Respuela P, Tebartz C, Kolovos P, Nikolic M, Fueyo R, et al. PRC2 facilitates the regulatory topology required for poised enhancer function during pluripotent stem cell differentiation. *Cell Stem Cell* 2017;20:689–705 e9.
- [49] Entrevan M, Schuettengruber B, Cavalli G. Regulation of genome architecture and function by polycomb proteins. *Trends Cell Biol* 2016;26:511–25.
- [50] Kwasniewski JC, Fiore C, Chaudhari HG, Cohen BA. High-throughput functional testing of ENCODE segmentation predictions. *Genome Res* 2014;24:1595–602.
- [51] Rajagopal N, Srinivasan S, Kooshesh K, Guo Y, Edwards MD, Banerjee B, et al. High-throughput mapping of regulatory DNA. *Nat Biotechnol* 2016;34:167–74.
- [52] Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. *Science* 2002;295:1306–11.
- [53] 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;326:289–93.
- [54] Fullwood MJ, Liu MH, Pan YF, Liu J, Xu H, Mohamed YB, et al. An oestrogen-receptor- α -bound human chromatin interactome. *Nature* 2009;462:58–64.
- [55] Mumbach MR, Rubin AJ, Flynn RA, Dai C, Khavari PA, Greenleaf WJ, et al. HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat Methods* 2016;13:919–22.
- [56] Forcato M, Nicoletti C, Pal K, Livi CM, Ferrari F, Bicciato S. Comparison of computational methods for Hi-C data analysis. *Nat Methods* 2017;14:679–85.
- [57] 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;174:744–57 e24.
- [58] Schoenfelder S, Fraser P. Long-range enhancer-promoter contacts in gene expression control. *Nat Rev Genet* 2019;20:437–55.
- [59] Robson MI, Ringel AR, Mundlos S. Regulatory landscaping: how enhancer-promoter communication is sculpted in 3D. *Mol Cell* 2019;74:1110–22.
- [60] Kempfer R, Pombo A. Methods for mapping 3D chromosome architecture. *Nat Rev Genet* 2019.
- [61] McCord RP, Kaplan N, Giorgetti L. Chromosome conformation capture and beyond: toward an integrative view of chromosome structure and function. *Mol Cell* 2020.
- [62] Lin D, Hong P, Zhang S, Xu W, Jamal M, Yan K, et al. Digestion-ligation-only Hi-C is an efficient and cost-effective method for chromosome conformation capture. *Nat Genet* 2018;50:754–63.
- [63] Liang Z, Li G, Wang Z, Djekidel MN, Li Y, Qian MP, et al. BL-Hi-C is an efficient and sensitive approach for capturing structural and regulatory chromatin interactions. *Nat Commun* 2017;8:1622.
- [64] Brant L, Georgomanolis T, Nikolic M, Brackley CA, Kolovos P, van Ijcken W, et al. Exploiting native forces to capture chromosome conformation in mammalian cell nuclei. *Mol Syst Biol* 2016;12:891.
- [65] Mizi A, Gade Gusmao E, Papantonis A. iHi-C 2.0: a simple approach for mapping native spatial chromatin organisation from low cell numbers. *Methods* 2020;170:33–7.
- [66] Beagrie RA, Scialdone A, Schueler M, Kraemer DC, Chotalia M, Xie SQ, et al. Complex multi-enhancer contacts captured by genome architecture mapping. *Nature* 2017;543:519–24.
- [67] Lai B, Tang Q, Jin W, Hu G, Wangsa D, Cui K, et al. Trac-looping measures genome structure and chromatin accessibility. *Nat Methods* 2018;15:741–7.
- [68] Redolfi J, Zhan Y, Valdes-Quezada C, Kryzhanovska M, Guerreiro I, Iesmantavicius V, et al. DamC reveals principles of chromatin folding in vivo without crosslinking and ligation. *Nat Struct Mol Biol* 2019;26:471–80.
- [69] Zheng M, Tian SZ, Capurso D, Kim M, Maurya R, Lee B, et al. Multiplex chromatin interactions with single-molecule precision. *Nature* 2019;566:558–62.
- [70] Li X, Luo OJ, Wang P, Zheng M, Wang D, Piecuch E, et al. Long-read ChIA-PET for base-pair-resolution mapping of haplotype-specific chromatin interactions. *Nat Protoc* 2017;12:899–915.
- [71] Tang Z, Luo OJ, Li X, Zheng M, Zhu JJ, Szalai P, et al. CTCF-mediated human 3D genome architecture reveals chromatin topology for transcription. *Cell* 2015;163:1611–27.
- [72] Rowley MJ, Nichols MH, Lyu X, Ando-Kuri M, Rivera ISM, Hermetz K, et al. Evolutionarily conserved principles predict 3D chromatin organization. *Mol Cell* 2017;67:837–52 e7.
- [73] Weintraub AS, Li CH, Zamudio AV, Sigova AA, Hannett NM, Day DS, et al. YY1 is a structural regulator of enhancer-promoter loops. *Cell* 2017;171:1573–88 e28.
- [74] Fang R, Yu M, Li G, Chee S, Liu T, Schmitt AD, et al. Mapping of long-range chromatin interactions by proximity ligation-assisted ChIP-seq. *Cell Res* 2016;26:1345–8.
- [75] Hughes JR, Roberts N, McGowan S, Hay D, Giannoulitou E, Lynch M, et al. Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nat Genet* 2014;46:205–12.
- [76] Oudelaar AM, Beagrie RA, Gosden M, De Ornellas S, Georgiades E, Kerry J, et al. Dissection of the 4D chromatin structure of the α -globin locus through in vivo erythroid differentiation with extreme spatial and temporal resolution. *bioRxiv* 2019:763763.
- [77] Javierre BM, Burren OS, Wilder SP, Kreuzhuber R, Hill SM, Sewitz S, et al. Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. *Cell* 2016;167:1369–84 e19.
- [78] Schoenfelder S, Furlan-Magaril M, Mifsud B, Tavares-Cadete F, Sugar R, Javierre B-M, et al. The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. *Genome Res* 2015;25:582–97.
- [79] Kolovos P, Brouwer RWW, Kockx CEM, Lesnussa M, Kepner N, Zuin J, et al. Investigation of the spatial structure and interactions of the genome at sub-kilobase-pair resolution using T2C. *Nat Protoc* 2018;13:459–77.
- [80] Sahlen P, Abdullayev I, Ramskold D, Matskova N, Rilakovic N, Lotstedt B, et al. Genome-wide mapping of promoter-anchored interactions with close to single-enhancer resolution. *Genome Biol* 2015;16:156.
- [81] Muller H, Scolari VF, Agier N, Piazza A, Thierry A, Mercy G, et al. Characterizing meiotic chromosomes' structure and pairing using a designer sequence optimized for Hi-C. *Mol Syst Biol* 2018;14:e8293.
- [82] Hsieh TH, Weiner A, Lajoie B, Dekker J, Friedman N, Rando OJ. Mapping nucleosome resolution chromosome folding in yeast by micro-C. *Cell* 2015;162:108–19.
- [83] Ma W, Ay F, Lee C, Gulsoy G, Deng X, Cook S, et al. Fine-scale chromatin interaction maps reveal the cis-regulatory landscape of human lincRNA genes. *Nat Methods* 2015;12:71–8.
- [84] Ben Zouari Y, Molitor AM, Sikorska N, Pancaldi V, Sexton T. ChiCMaxima: a robust and simple pipeline for detection and visualization of chromatin looping in Capture Hi-C. *Genome Biol* 2019;20:102.

- [85] Mateo LJ, Murphy SE, Hafner A, Cinquini IS, Walker CA, Boettiger AN. Visualizing DNA folding and RNA in embryos at single-cell resolution. *Nature* 2019;568:49–54.
- [86] Kaul A, Bhattacharyya S, Ay F. Identifying statistically significant chromatin contacts from Hi-C data with FitHiC2. *Nat Protoc* 2020.
- [87] Oudelaar AM, Davies JOJ, Hanssen LLP, Telenius JM, Schwessinger R, Liu Y, et al. Single-allele chromatin interactions identify regulatory hubs in dynamic compartmentalized domains. *Nat Genet* 2018;50:1744–51.
- [88] Sutherland H, Bickmore WA. Transcription factories: gene expression in unions? *Nat Rev Genet* 2009;10:457–66.
- [89] Jiang T, Raviram R, Snetkova V, Rocha PP, Proudhon C, Badri S, et al. Identification of multi-loci hubs from 4C-seq demonstrates the functional importance of simultaneous interactions. *Nucleic Acids Res* 2016;44:8714–25.
- [90] Ay F, Vu TH, Zeitzi MJ, Varoquaux N, Carette JE, Vert JP, et al. Identifying multi-locus chromatin contacts in human cells using tethered multiple 3C. *BMC Genomics* 2015;16:121.
- [91] Olivares-Chauvet P, Mukamel Z, Lifshitz A, Schwartzman O, Elkayam NO, Lubling Y, et al. Capturing pairwise and multi-way chromosomal conformations using chromosomal walks. *Nature* 2016;540:296–300.
- [92] Allahyar A, Vermeulen C, Bouwman BAM, Krijger PHL, Verstegen M, Geeven G, et al. Enhancer hubs and loop collisions identified from single-allele topologies. *Nat Genet* 2018;50:1151–60.
- [93] Vermeulen C, Allahyar A, Bouwman BAM, Krijger PHL, Verstegen M, Geeven G, et al. Multi-contact 4C: long-molecule sequencing of complex proximity ligation products to uncover local cooperative and competitive chromatin topologies. *Nat Protoc* 2020;15:364–97.
- [94] Nagano T, Lubling Y, Stevens TJ, Schoenfelder S, Yaffe E, Dean W, et al. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* 2013;502:59–64.
- [95] Flyamer IM, Gassler J, Imakaev M, Brandao HB, Ulianov SV, Abdennur N, et al. Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature* 2017;544:110–4.
- [96] Ramani V, Deng X, Qiu R, Gunderson KL, Steemers FJ, Distech CM, et al. Massively multiplex single-cell Hi-C. *Nat Methods* 2017;14:263–6.
- [97] Du Z, Zheng H, Huang B, Ma R, Wu J, Zhang X, et al. Allelic reprogramming of 3D chromatin architecture during early mammalian development. *Nature* 2017;547:232–5.
- [98] Stevens TJ, Lando D, Basu S, Atkinson LP, Cao Y, Lee SF, et al. 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature* 2017;544:59–64.
- [99] Tan L, Xing D, Chang CH, Li H, Xie XS. Three-dimensional genome structures of single diploid human cells. *Science* 2018;361:924–8.
- [100] Wang Y, Song F, Zhang B, Zhang L, Xu J, Kuang D, et al. The 3D Genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. *Genome Biol* 2018;19:151.
- [101] Huang D, Yi X, Zhang S, Zheng Z, Wang P, Xuan C, et al. GWAS4D: multidimensional analysis of context-specific regulatory variant for human complex diseases and traits. *Nucleic Acids Res* 2018;46:W114–20.
- [102] Hariprakash JM, Ferrari F. Computational Biology Solutions to Identify Enhancers-target Gene Pairs. *Comput Struct Biotechnol J* 2019;17:821–31.
- [103] Cao Q, Yip KY. A survey on computational methods for enhancer and enhancer target predictions. *Conf Proc* 2015.
- [104] Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, et al. The accessible chromatin landscape of the human genome. *Nature* 2012;489:75–82.
- [105] Shlyueva D, Stampfel G, Stark A. Transcriptional enhancers: from properties to genome-wide predictions. *Nat Rev Genet* 2014;15:272–86.
- [106] Corradin O, Saikhova A, Akhtar-Zaidi B, Myeroff L, Willis J, Cowper-Salari R, et al. Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits. *Genome Res* 2014;24:1–13.
- [107] 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;485:376–80.
- [108] Hnisz D, Day DS, Young RA. Insulated neighborhoods: structural and functional units of mammalian gene control. *Cell* 2016;167:1188–200.
- [109] Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, et al. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* 2009;459:108–12.
- [110] Fulco CP, Nasser J, Jones TR, Munson G, Bergman DT, Subramanian V, et al. Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. *Nat Genet* 2019;51:1664–9.
- [111] 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. *Mol Cell* 2018;71:858–71 e8.
- [112] Mehdi T, Bailey SD, Guilhamon P, Lupien M. C3D: a tool to predict 3D genomic interactions between cis-regulatory elements. *Bioinformatics* 2019;35:877–9.
- [113] Kodzius R, Kojima M, Nishiyori H, Nakamura M, Fukuda S, Tagami M, et al. CAGE: cap analysis of gene expression. *Nat Methods* 2006;3:211–22.
- [114] Yao L, Shen H, Laird PW, Farnham PJ, Berman BP. Inferring regulatory element landscapes and transcription factor networks from cancer methylomes. *Genome Biol* 2015;16:105.
- [115] Lu Y, Zhou Y, Tian W. Combining Hi-C data with phylogenetic correlation to predict the target genes of distal regulatory elements in human genome. *Nucleic Acids Res* 2013;41:10391–402.
- [116] Naville M, Ishibashi M, Ferg M, Bengani H, Rinkwitz S, Krecsmarik M, et al. Long-range evolutionary constraints reveal cis-regulatory interactions on the human X chromosome. *Nat Commun* 2015;6:6904.
- [117] Clement Y, Torbey P, Gilardi-Hebenstreit P, Crollius HR. Enhancer-gene maps in the human and zebrafish genomes using evolutionary linkage conservation. *Nucleic Acids Res* 2020.
- [118] Friedman J, Hastie T, Tibshirani R. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* 2008;9:432–41.
- [119] Zhu Y, Chen Z, Zhang K, Wang M, Medovoy D, Whitaker JW, et al. Constructing 3D interaction maps from 1D epigenomes. *Nat Commun* 2016;7:10812.
- [120] Liu D, Davila-Velderrain J, Zhang Z, Kellis M. Integrative construction of regulatory region networks in 127 human reference epigenomes by matrix factorization. *Nucleic Acids Res* 2019;47:7235–46.
- [121] Rennie S, Dalby M, van Duin L, Andersson R. Transcriptional decomposition reveals active chromatin architectures and cell specific regulatory interactions. *Nat Commun* 2018;9:487.
- [122] Dryden NH, Broome LR, Dudbridge F, Johnson N, Orr N, Schoenfelder S, et al. Unbiased analysis of potential targets of breast cancer susceptibility loci by Capture Hi-C. *Genome Res* 2014;24:1854–68.
- [123] He B, Chen C, Teng L, Tan K. Global view of enhancer-promoter interactome in human cells. *Proc Natl Acad Sci U S A* 2014;111:E2191–9.
- [124] Zhao C, Li X, Hu H. PETModule: a motif module based approach for enhancer target gene prediction. *Sci Rep* 2016;6:30043.
- [125] Roy S, Shahrirani AF, Chasman D, Knaack S, Ay F, Stewart R, et al. A predictive modeling approach for cell line-specific long-range regulatory interactions. *Nucleic Acids Res* 2015;43:8694–712.
- [126] Whalen S, Truty RM, Pollard KS. Enhancer-promoter interactions are encoded by complex genomic signatures on looping chromatin. *Nat Genet* 2016;48:488–96.
- [127] Al Bkhetan Z, Plewczynski D. Three-dimensional epigenome statistical model: genome-wide chromatin looping prediction. *Sci Rep* 2018;8:5217.
- [128] Talukder A, Saadat S, Li X, Hu H. EPIP: a novel approach for condition-specific enhancer-promoter interaction prediction. *Bioinformatics* 2019.
- [129] Singh S, Yang Y, Poczos B, Ma J. Predicting enhancer-promoter interaction from genomic sequence with deep. *Neural Networks* 2018;:085241.
- [130] Mao W, Kostka D, Chikina M. Modeling enhancer-promoter interactions with attention-based. *Neural Networks* 2017;:219667.
- [131] Yang Y, Zhang R, Singh S, Ma J. Exploiting sequence-based features for predicting enhancer-promoter interactions. *Bioinformatics* 2017;33:i252–60.
- [132] Zeng W, Wu M, Jiang R. Prediction of enhancer-promoter interactions via natural language processing. *BMC Genomics* 2018;19:84.
- [133] Cao F, Zhang Y, Loh YP, Cai Y, Fullwood MJ. Predicting chromatin interactions between open chromatin regions from DNA sequences. 2019:720748.
- [134] Schuster M, Paliwal KK. Bidirectional recurrent neural networks. *IEEE Trans Signal Process* 1997;45:2673–81.
- [135] Hafez D, Karabacak A, Krueger S, Hwang YC, Wang LS, Zinzen RP, et al. McEnhancer: predicting gene expression via semi-supervised assignment of enhancers to target genes. *Genome Biol* 2017;18:199.
- [136] Schreiber J, Libbrecht M, Birmes J, Noble W. Nucleotide sequence and DNaseI sensitivity are predictive of 3D chromatin architecture. 2018:103614.
- [137] Li W, Wong WH, Jiang R. DeepTACT: predicting 3D chromatin contacts via bootstrapping deep learning. *Nucleic Acids Res* 2019.
- [138] Zhang H, Li F, Jia Y, Xu B, Zhang Y, Li X, et al. Characteristic arrangement of nucleosomes is predictive of chromatin interactions at kilobase resolution. *Nucleic Acids Res* 2017;45:12739–51.
- [139] Ghandi M, Lee D, Mohammad-Noori M, Beer MA. Enhanced regulatory sequence prediction using gapped k-mer features. *PLoS Comput Biol* 2014;10:e1003711.
- [140] Kelley DR, Snook J, Rinn JL. Basset: learning the regulatory code of the accessible genome with deep convolutional neural networks. *Genome Res* 2016;26:990–9.
- [141] Cao Q, Anyansi C, Hu X, Xu L, Xiong L, Tang W, et al. Reconstruction of enhancer-target networks in 935 samples of human primary cells, tissues and cell lines. *Nat Genet* 2017;49:1428–36.
- [142] Hait TA, Amar D, Shamir R, Elkon R. FOCUS: a novel method for analyzing enhancer and gene activity patterns infers an extensive enhancer-promoter map. *Genome Biol* 2018;19:56.
- [143] Hait TA, Elkon R, Shamir R. CT-FOCUS: a novel method for inferring cell type-specific enhancer-promoter maps. *bioRxiv* 2019:707158.
- [144] Moore JE, Pratt HE, Purcaro MJ, Weng Z. A curated benchmark of enhancer-gene interactions for evaluating enhancer-target gene prediction methods. *Genome Biol* 2020;21:17.
- [145] Cao F, Fullwood MJ. Inflated performance measures in enhancer-promoter interaction-prediction methods. *Nat Genet* 2019;51:1196–8.
- [146] Belokopytova PS, Nuriddinov MA, Mozheiko EA, Fishman D, Fishman V. Quantitative prediction of enhancer-promoter interactions. *Genome Res* 2020;30:72–84.
- [147] Lundberg SM, Erion GG, Lee S-I. Consistent individualized feature attribution for tree ensembles. *arXiv e-prints* 2018.
- [148] Shrikumar A, Greenside P, Shcherbina A, Kundaje A. Not just a black box: learning important features through propagating activation differences. *arXiv e-prints* 2016.

- [149] Li Y, Chen CY, Wasserman WW. Deep feature selection: theory and application to identify enhancers and promoters. *J Comput Biol* 2016;23:322–36.
- [150] Zhang S, Chasman D, Knaack S, Roy S. In silico prediction of high-resolution Hi-C interaction matrices. *Nat Commun* 2019;10:5449.
- [151] Di Piero M, Cheng RR, Lieberman Aiden E, Wolynes PG, Onuchic JN. De novo prediction of human chromosome structures: epigenetic marking patterns encode genome architecture. *Proc Natl Acad Sci U S A* 2017;114:12126–31.
- [152] Qi Y, Zhang B. Predicting three-dimensional genome organization with chromatin states. *PLoS Comput Biol* 2019;15:e1007024.
- [153] Kim M, Zheng M, Tian SZ, Lee B, Chuang JH, Ruan Y. MIA-Sig: multiplex chromatin interaction analysis by signal processing and statistical algorithms. *Genome Biol* 2019;20:251.
- [154] Zhang R, Ma J. Probing multi-way chromatin interaction with hypergraph representation learning. 2020:2020.01.22.916171.
- [155] Visel A, Minovitsky S, Dubchak I, Pennacchio LA. VISTA enhancer browser—a database of tissue-specific human enhancers. *Nucleic Acids Res* 2007;35: D88–92.
- [156] Patwardhan RP, Hiatt JB, Witten DM, Kim MJ, Smith RP, May D, et al. Massively parallel functional dissection of mammalian enhancers in vivo. *Nat Biotechnol* 2012;30:265–70.
- [157] Boettiger A, Murphy S. Advances in chromatin imaging at kilobase-scale resolution. *Trends Genet* 2020.
- [158] Morgan SL, Mariano NC, Bermudez A, Arruda NL, Wu F, Luo Y, et al. Manipulation of nuclear architecture through CRISPR-mediated chromosomal looping. *Nat Commun* 2017;8:15993.
- [159] Liu X, Zhang Y, Chen Y, Li M, Zhou F, Li K, et al. In situ capture of chromatin interactions by biotinylated dCas9. *Cell* 2017;170. 1028–43 e19.
- [160] Wang H, Xu X, Nguyen CM, Liu Y, Gao Y, Lin X, et al. CRISPR-mediated programmable 3D genome positioning and nuclear organization. *Cell* 2018;175. 1405–17 e14.
- [161] Kim JH, Rege M, Valeri J, Dunagin MC, Metzger A, Titus KR, et al. LADL: light-activated dynamic looping for endogenous gene expression control. *Nat Methods* 2019;16:633–9.
- [162] Fishilevich S, Nudel R, Rappaport N, Hadar R, Plaschkes I, Iny Stein T, et al. GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. Database (Oxford) 2017;2017.
- [163] Duggal G, Wang H, Kingsford C. Higher-order chromatin domains link eQTLs with the expression of far-away genes. *Nucleic Acids Res* 2014;42:87–96.
- [164] Gate RE, Cheng CS, Aiden AP, Siba A, Tabaka M, Lituiev D, et al. Genetic determinants of co-accessible chromatin regions in activated T cells across humans. *Nat Genet* 2018;50:1140–50.
- [165] Chandra V, Bhattacharyya S, Schmiedel BJ, Madrigal A, Fotsing S, Seumois G, et al. Promoter-interacting expression quantitative trait loci (pieQTLs) in human immune cell types. SSRN. 2019; <http://dx.doi.org/10.2139/ssrn.3402070>.
- [166] Zheng Z, Huang D, Wang J, Zhao K, Zhou Y, Guo Z, et al. QTLbase: an integrative resource for quantitative trait loci across multiple human molecular phenotypes. *Nucleic Acids Res* 2020;48:D983–91.
- [167] Sadowski M, Kraft A, Szalaj P, Wlasnowolski M, Tang Z, Ruan Y, et al. Spatial chromatin architecture alteration by structural variations in human genomes at the population scale. *Genome Biol* 2019;20:148.
- [168] Xie S, Duan J, Li B, Zhou P, Hon GC. Multiplexed engineering and analysis of combinatorial enhancer activity in single cells. *Mol Cell* 2017;66. 285–99 e5.
- [169] Gasperini M, Hill AJ, McFaline-Figueroa JL, Martin B, Kim S, Zhang MD, et al. A genome-wide framework for mapping gene regulation via cellular genetic screens. *Cell* 2019;176. 377–90 e19.