COCOA: A Framework for Fine-scale Mapping Cell-type-specific Chromatin Compartmentalization Using Epigenomic Information

Kai Li[1, #], Ping Zhang[1, #], Jinsheng Xu[1], Zi Wen[1], Junying Zhang[1], Zhike Zi[3, *] and Li Li[1, 2, *]

[1] Hubei Key Laboratory of Agricultural Bioinformatics, College of Informatics, Huazhong Agricultural University, Wuhan 430070, China

[2] Hubei Hongshan Laboratory, Huazhong Agricultural University, Wuhan 430070, China

[3] Shenzhen Key Laboratory of Synthetic Genomics, Guangdong Provincial Key Laboratory of Synthetic Genomics, Key Laboratory of Quantitative Synthetic Biology, Shenzhen Institute of Synthetic Biology, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China

[*] Corresponding author:

Li Li, College of Informatics, Huazhong Agricultural University, Wuhan 430070, China.

Email: li.li@mail.hzau.edu.cn

Zhike Zi, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China

Email: zk.zi@siat.ac.cn

[#]Equal contribution.

Abstract

Chromatin compartmentalization and epigenomic modification are crucial factors in cell differentiation and diseases development. However, mapping precise chromatin compartmental patterns across multiple cell types requires Hi-C or Micro-C data at high sequencing depth. Exploring the systematic relationship between epigenomic modifications and compartmental patterns remains a challenge. To address these issues, we present COCOA, a deep neural network framework that uses convolution and attention mechanisms to infer reliable fine-scale chromatin compartment patterns from six representative histone modification signals. COCOA achieves this by extracting 1-D track features through bi-directional feature reconstruction after resolution-specific binning epigenomic signals. These track features are then cross-fused with contact features using an attention mechanism. Subsequently, the contact features are transformed into chromatin compartment patterns through residual feature reduction. COCOA demonstrates accurate inference of chromatin compartmentalization at a fine-scale resolution and exhibits stable performance on test sets. In addition, we explored the impact of histone modifications on the chromatin compartmentalization through *in silico* epigenomic perturbation experiments. When using 1kb resolution high-depth experimental data, obscure compartments are observed, whereas COCOA can generate clear and detailed compartmental patterns. Finally, we demonstrated that COCOA enables cell-type-specific prediction of unrevealed chromatin compartment patterns in various biological processes. Thus, COCOA is an effective tool for gaining chromatin compartmentalization insights from epigenomics in a wide range of biological scenarios.

KEYWORDS: 3D genome; Deep learning; Chromatin compartment; Histone modification; Cell-type-specificity

Introduction

The three-dimensional (3D) architecture of chromatin is essential for gene expression during cell differentiation and disease development [1,2]. Recent advances in next generation sequencing have led to the development of several chromosome conformation capture techniques, such as Hi-C, Micro-C, and Pore-C [3–5], enabling the exploration of multiscale chromatin structural elements including chromatin compartment [3,6], topological associating domains (TADs) [7,8], loops [6], stripes [9] and microcompartments [10]. These techniques have revealed that the chromatin can be segregated into A compartment and B compartment [3,11]. The A compartments are generally active in the genome, whereas the B compartments are mostly transcriptional repressive. These chromatin compartments are closely related to the mechanisms underlying various key biological processes [12,13].

To identify chromatin compartments, sequencing data is usually processed into contact maps and distance effects are eliminated using normalization methods. The normalized contact map is then used to calculate the correlation matrix (CM), which is subjected to principal component analysis (PCA). The sign of the first principal component (PC1) corresponds to the compartment state [3]. Most analyses related to chromatin compartments rely on CM and PC1 [14–16]. While the CM is commonly available and of high-quality at mega-base scale, it becomes noisy for resolution finer than 25kb, failed to show clear plaid patterns due to its sparseness. Recent studies have suggested the associations of the fine-scale resolution chromatin compartments with other structural elements [17,18], histone modifications and chromatin accessibility [19]. However, the available chromatin compartment data does not match the scale of the epigenomic data, making the connection between epigenomics and chromatin compartmentalization a challenge. Furthermore, due to technical limitations and sequencing costs [20], experimentally mapping high-resolution chromatin compartment is both expensive and labour-intensive. Therefore, there is unmet need for the development of a computational method to obtain the fine-scale CM across multiple cell lines.

In the past decade, deep learning [21] has emerged as a widely used tool in computational 3D genomes. These applications include various tasks such as TAD boundary recognition [22,23], chromatin loop detection [24,25], chromatin interaction data enhancement [26,27], interaction matrix generation [28–30] and single cell Hi-C imputation [31,32]. While several methods explore contact map generation and enhancement, they lack cell type specificity. For example, HiC-Reg [33] used fourteen epigenomic signals from five cell lines to predict short-range chromatin interactions using random forests. Akita [29] and Orca [34] adopted the convolutional neural network to predict contact maps from DNA sequences. However, these methods are not capable directly inferring contact maps across different cell types. Recently, two proposed methods, C.Origami [35] and Epiphany [30], address this limitation by utilizing the histone modifications and chromatin accessibility data. C.Origami predicts short-range interactions by integrating CTCF, chromatin accessibility, and DNA sequence information through a neural network containing the attention and convolutional modules. Epiphany uses multiple epigenomic signals to generate short-range chromatin contact maps. However, these methods have their own limitations in terms of chromatin compartmentation and method generalization. Firstly, these existing methods concentrate on the prediction of short-range interactions (TADs and loops) while ignoring long-range interactions (compartments). Additionally, the relationship between compartmentalization and histone modification is still unresolved.

Furthermore, these models require inputs in fixed bin sizes, limiting scalability and preventing across-resolution predictions.

To address these limitations, we introduce COCOA (mapping **chro**matin **co**mpartmentalization with epigenomic inform**a**tion), a method that predict the cell-type-specific CM using six types of accessible epigenomic modification signals. COCOA adopts bidirectional feature reconstruction and cross attention fusion for bi-directional reconstruction and fusion of epigenomics data. Subsequently, residual feature reduction is applied to map the fused results into CM. COCOA is specifically designed to generate chromatin compartmentalization, and the predicted CM can be directly used to determine compartment status. We evaluated the performance of COCOA using multiple metrics (MSE, MAE, GenomeDISCO, PCC, SSIM and PSNR). The results demonstrate that COCOA accurately generates significant and biologically meaningful CM. Furthermore, we conducted *in silico* perturbation experiments to investigate the influence of histone modifications on compartments. Additionally, we tested the generalization performance of COCOA by making model predictions with resolution-specific and cell-type-specific data. The results show that COCOA enables robust performance at various resolutions from diverse cell lines, providing insights into the patterns of chromatin compartments in immune and disease tissues.

Method

Hi-C and Micro-C data sources and pre-processing

We collected publicly available processed Hi-C and Micro-C data of different cell lines from the 4DN database [36]. Intra-chromosomal contact maps were computed from these data for model training and testing (see details in Table S1). Depending on the specific task, the intra-chromosomal contact maps is computed at different resolutions using the cooler package [37]. To eliminate the distance effect in the contact maps, we applied the observed-expected normalization method [3]. Finally, these normalized contact maps were converted into CMs, which clearly depict plaid pattern of chromatin compartmentalization.

ChIP-seq data sources and pre-processing

Histone modification signals (H3K9me3, H3K27ac, H3K4me1, H3K27me3, H3K4me3, H3K36me3) from the ChIP-seq [38] data for all cells were retrieved from ENCODE project [39] (see details in Table S2). The ChIP-seq data were binned to specific resolutions using the pyBigWig package (Figure 1A). After binning, a $\log(x + 1)$ transformation and min-max normalization were performed on the data. Finally, the processed data were combined into an epigenomic signal matrix.

Dividing matrices

The pre-processing step generates two matrices: a symmetric correlation matrix (CM) with dimensions n×n ($CM_{n \times n}$) and an epigenomic matrix (EM) with dimensions m×n ($EM_{m \times n}$). Each value CM_{ij} in the $CM_{n \times n}$ represents the correlation strength between genomic segment i and j. Values greater than one indicate that the two genomic segments have the same interaction mode, while less than one indicate the opposite interaction mode,

providing information about the status of chromatin compartments. Each value EM_{ij} in the $EM_{m \times n}$ represents the signal strength of genomic segment j on epigenomic track i.

To better preserve the plaid pattern of chromatin compartmentalization and adapt to the inputs of neural network, we implemented the following processing scheme. First, the $CM_{n \times n}$ was divided into sub-matrices of $k \times k$ size ($SCM_{k \times k}$), and the $EM_{m \times n}$ were divided into two sub-matrices of $m \times k$ size ($SLEM_{m \times k}$ and $SREM_{m \times k}$). We started at the diagonal position in the top-left corner of $CM_{n \times n}$ and moved horizontally, dividing it into $SCM_{k \times k}$. Simultaneously, we divided the two corresponding groups of genomic loci from EM into $SLEM_{m \times k}$ and $SREM_{m \times k}$. After completing the horizontal dividing, we moved the current position diagonally by k positions. This process was repeated until the entire $CM_{n \times n}$ could no longer be divided. Due to computational resource constraints, we sampled the $SCM_{k \times k}$ in groups to minimize the size of the training datasets ($SLEM_{m \times k}$ and $SREM_{m \times k}$ were synchronized to minimize). Finally, these data were saved separately for further modelling.

Combining predicted sub-matrices

The COCOA model takes the $SLEM_{m \times k}$ and $SREM_{m \times k}$ for each chromosome division as inputs. It then outputs a series of predicted correlation sub-matrices. These sub-matrices sequentially covered a square matrix ($PCM_{n \times n}$) with the same number of columns as the $EM_{m \times n}$. The specific coordinates for covering each predicted correlation sub-matrices are determined by the corresponding inputs ($SLEM_{m \times k}$ and $SREM_{m \times k}$). Finally, the complete $PCM_{n \times n}$ is generated and saved for further biological analyses.

COCOA architecture

The COCOA model consists of three main components: bidirectional feature reconstruction, cross attention fusion, and residual feature reduction (Figure 1B), which are described in the following sections.

Bidirectional feature reconstruction

The bidirectional feature reconstruction module consists of two matrix reconstruction layers (MR layers). The construction of these MR layers is inspired by our previous work on chromatin interaction data enhancement [40]. Each MR layer consists of two parts: an aggregation convolution layer with a filter size of N × 1 and a linear reconstruction layer. The output of each MR layer is computed as follows:

$$v(SEM) = Tanh\big((SEM \otimes K_{N \times 1}) \times W_i\big) \tag{1}$$

$$MR(SEM) = v(SEM) \cdot v(SEM)^T \times W_j \tag{2}$$

where $\otimes$ denotes the convolution operation, $K_{N \times 1}$ represents convolution kernel (N × 1), Tanh is the activation function [41], × denotes Hadamard product and · denotes dot product. SEM represents $SLEM_{m \times k}$ or $SREM_{m \times k}$ generated through pre-processing. v^T represents the transposition of the vector v. W_i and W_j are learnable weight matrices, respectively. The MR layer aggregates multiple ChIP-seq track signals from different genomic loci into a 1D vector. This vector is in turn reconstructed into a low-ranking epigenomic track features using learnable

weight matrices. In summary, this module obtains bi-directional epigenomic track features by reconstructing the $SLEM_{m \times k}$ and $SREM_{m \times k}$.

Cross attention fusion

Next, the COCOA model employs the cross attention fusion module to fuse bi-directional epigenomic track features. The cross attention fusion module mainly contains two attention feature fusion layers (AFF layers) [42]. Each AFF layer has three parts: global feature extraction, local feature extraction and attention fusion. The results of cross attention fusion are defined as follows:

$$CAF\ (P, Q) = \ concat(AFF_1(P, Q), AFF_2(Q, P)) \qquad (3)$$

where P and Q represent bi-directional epigenomic track features, respectively. concat denotes stacking two outputs in the same dimension. AFF refers to an attention-based uniform and general neural network layer for feature fusion proposed by Dai et al. [42]. The cross attention fusion module transforms the epigenomic track features in the other direction into potential attention weights to reinforce the epigenomic track features in the current direction. By interleaving attention fusion and concatenation, a set of fused contact feature maps is obtained as inputs for the next module.

Residual feature reduction

The residual feature reduction module consists of a series of residual blocks, each containing several residual layers. Following the approach described in previous work [43], each residual layer is composed of convolutional layers with different convolution kernels, BN layers [44], and activation functions. The computation of each layer is defined as follows:

$$F(X) = Tanh\big(BN_1(X \otimes K_{n \times n})\big) \qquad (4)$$

$$Res(X) = Tanh\big(\ BN_2(F(X) \otimes K_{m \times m}) + X\big) \qquad (5)$$

where K denotes the convolution kernels of different size, Tanh is activation function and BN is the BatchNorm layer. X represents the fused contact feature maps for the first layer and the output of the current layer serving as the input for the next layer. The residual feature reduction module decreases the channels of the contact features from the previous module, level by level. Throughout this process, the residual layer continuously filters to retain important information from the previous layer, aggregating it with the output of current layer. Finally, the predicted correlation sub-matrix is obtained from the last layer of residual feature reduction module.

Loss function

The COCOA model can be seen as a function F with a parameter set θ, which maps each group input $SLEM_{i,m \times k}$ and $SREM_{i,m \times k}$ to the predicted correlation sub-matrix $PSCM_{i,k \times k}$ (i.e., $PSCM_{i,k \times k} = F(SLEM_{i,m \times k}, SREM_{i,m \times k} : \theta)$). The objective of the training is to find a set of θ^* to enable $PSCM_{i,k \times k}$ similar to the ground truth $SCM_{i,k \times k}$. Therefore, COCOA initially uses the mean square error (MSE) loss to minimize the pairwise error of genomic range $k \times k$ between PSCM and SCM. This loss can be described as:

$$\mathfrak{L}_{MSE}\left(PSCM_{i,k\times k}, SCM_{i,k\times k}\right) = \frac{1}{m}\sum_{i=1}^{m}\left(PSCM_{i,k\times k} - SCM_{i,k\times k}\right)^2 \qquad (6)$$

Subsequently, COCOA incorporates a perceptual loss based on the VGG network [45] to restore structural information of the correlation matrix. Furthermore, we added the total variation (TV) loss [46], which effectively smooths noise in computer vision, as a regularization term to suppress the noise of the $PSCM_{k\times k}$. These losses can be described as:

$$\mathfrak{H}_{VGG}\left(PSCM_{i,k\times k}, SCM_{i,k\times k}\right) = \frac{1}{N}\sum_{k=1}^{N}\left(VGG\left(PSCM_{i,k\times k}\right) - VGG\left(SCM_{i,k\times k}\right)\right)^2 \qquad (7)$$

$$\mathfrak{F}_{TV}(PSCM_{k\times k}) = \sum_{i,j}\left(PSCM_{i,j-1} - PSCM_{i,j}\right)^2 + \sum_{i,j}\left(PSCM_{i+1,j} - PSCM_{i,j}\right)^2 \qquad (8)$$

Finally, the training objective can be represented as:

$$\theta^* = argmin_\theta\left[\mathfrak{L}_{MSE}\left(PSCM_{i,k\times k}, SCM_{i,k\times k}\right) + \alpha * \mathfrak{F}_{TV}\left(PSCM_{i,k\times k}\right) + \beta * \mathfrak{H}_{VGG}\left(PSCM_{i,k\times k}, SCM_{i,k\times k}\right)\right] \quad (9)$$

where α,β are scaling weights that range from 0 to 1.

COCOA training and hyper-parameter exploration

Before model training, we pre-processed each chromosome of the HFFc6 Micro-C data [47] and corresponding ChIP-seq data. Chromosome 1, 3, 5, 7, 9, 11, 13, 15, 17 and 19 were used as training sets, while chromosomes 18, 20, 21 and 22 were utilized for hyper-parameter tuning. The remaining chromosomes were allocated for performance evaluation.

The COCOA model was implemented on Python 3.7 with PyTorch1.12 [48]. We trained the model with a batch size of 16 for 120 epochs, using the Adam optimizer [49] with an initial learning rate of 5e-4 ($lr_{init} = 5e-4$). The details on model training and hyper-parameters are provided in Note S1.

Model evaluation

We started the evaluation process by making predictions on independent test sets using the best-trained model. The predicted correlation sub-matrices were then combined to form the intra-chromatin correlation matrix. The experimental 25 kb chromatin interaction correlation matrices were considered as the ground truth. During the evaluation, we used the principal component analysis (PCA) provided by the sklearn package [50] to calculate the first principal component (PC1) of the two correlation matrices. PC1 is generally considered to represent the A/B compartment information. Additionally, we discretized PC1 to obtain the chromatin compartment state, which was saved separately.

To assess the model performance, we used several indictors including Mean Square Error (MSE, see Equation (6)), Mean Absolute Error (MAE), Structure Similarity Index (SSIM) (assessing the similarity of two correlation matrices) and Peak Signal to Noise Ratio (PSNR) (measuring the quality score of the correlation matrices) [51]. MAE, SSIM and PSNR are defined by the following equations:

$$MAE(\hat{Y}, Y) = \frac{1}{n} \sum_{i=1}^{n} |\hat{y}_i - y_i| \tag{10}$$

$$SSIM(\hat{Y}, Y) = \frac{(2\mu_{\hat{Y}}\mu_Y + C_1) * (2\sigma_{\hat{Y}Y} + C_2)}{(\mu_{\hat{Y}}^2 + \mu_Y^2 + C_1) * (\sigma_{\hat{Y}}^2 + \sigma_Y^2 + C_2)} \tag{11}$$

$$PSNR(\hat{Y}, Y) = 10 * \log_{10}\left(\frac{N}{MSE(\hat{Y}, Y)}\right) \tag{12}$$

where $\hat{Y}$ denotes the predicted correlation matrices and Y represents the real correlation matrices. Furthermore, considering the chromatin compartmentalization information of the correlation matrices, we evaluated their reproducibility using multiple Pearson Correlation Coefficient (PCC) and GenomeDISCO score [52].

Results

Overview of COCOA

In this study, we proposed COCOA as a method for accurately predicting cell-type-specific chromatin compartment patterns at fine-scale resolution by integrating epigenomic modification signals. COCOA only requires six epigenomic track signals as inputs, which are accessible for most tissues and cell lines in the ENCODE database [53]. The targets of COCOA are defined as the correlation matrices (CM) of OE-normalized contact maps, allowing for the maximum retention of chromatin compartment pattern information. The COCOA framework connects these inputs and targets through binning, prediction and combination operations (Figure 1A). Notably, in the binning process, we utilized the resolution-specific binning approach (i.e., $Bin_{epi} = Bin_{corr}$) instead of a single bin from each genome site approach (i.e., $Bin_{epi} = Bin_{corr} * resolution$). This choice greatly improves the practicality of COCOA.
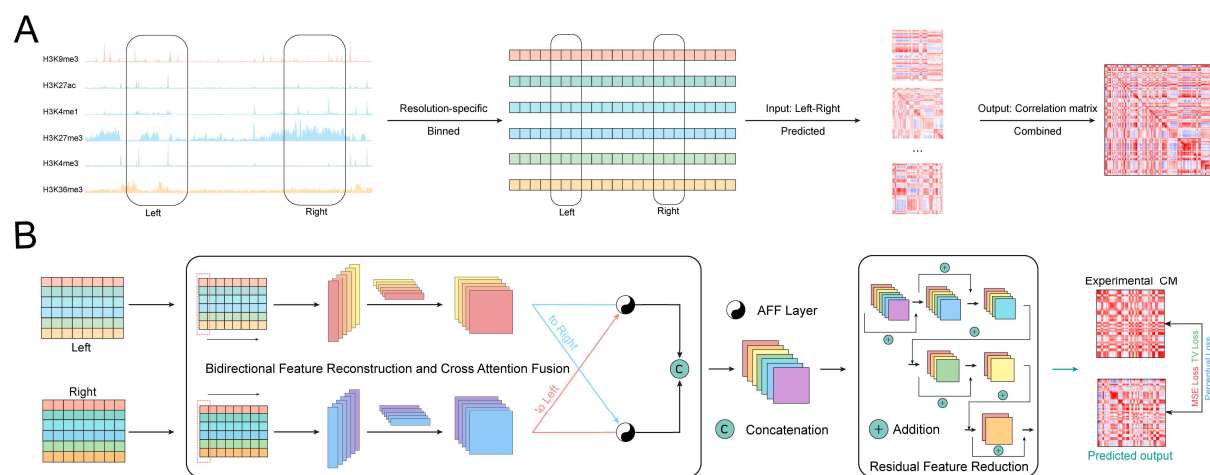
Figure 1 COCOA pipeline and architecture

(A) COCOA pipeline: The integration of six accessible epigenomic signals by resolution-specific binning serves as inputs to predict the correlation matrices.

**(B)** COCOA architecture: COCOA extracts 1D track features from each input (bidirectional feature reconstruction module) and then combines these features with spatial contact features (cross attention fusion module). The contact features are further processed by the residual feature reduction module to obtain the final prediction result. Parameters are updated using a backpropagation algorithm with mixed loss functions. (Refer to Method for detailed information)

We trained COCOA on Micro-C data of HFFc6 along with corresponding ChIP-seq data (Table S1 and S2) using backpropagation algorithm. Specifically, COCOA first utilizes the bidirectional feature reconstruction module to calculate the 1D track features separately from two inputs. This step captures the intrinsic association present in the original epigenomic data in each direction (Figure 1B and Method). Subsequently, the cross attention fusion module integrates these 1D track feature maps with space contact features based on crossed attention mechanisms (see details in Method). Lastly, the residual feature reduction module decodes these contact features to generate predicted results, which are then combined into a complete CM (Refer to the Combining predicted sub-matrices section of Method for detailed information). In addition, a composite loss function is employed to minimize the distance between the predicted targets and the ground truth.

COCOA accurately predicts chromatin compartmentalization pattern

To assess the performance of COCOA, we applied the trained COCOA model to randomly selected epigenomic data in test sets (Chr 12, Chr 14 and Chr 16) to generate predicted CM. We considered the CM calculated from the Micro-C data and its PC1 as the experimental CM, which can be regarded as the ground truth for comparison. Heat maps in Figure 2A and Figure S1A compare typical genomic regions using the predicted and the experimental CM. The results demonstrate that the predicted CM generally exhibit the correct chromatin compartmentalization pattern. Furthermore, COCOA shows outstanding generative capacity in capturing subtle chromatin compartments. Notably, the predicted CM show more pronounced interactions in dissimilar chromatin compartment blocks (blue blocks) compared with the experimental CM, while exhibiting partial over-reinforcement in identical chromatin compartment blocks (red blocks). We also computed the PC1 of the predicted CM and the experimental CM using the sklearn package [50]. Subsequently, the CMs were sorted based on the size of their respective PC1 values (Figure 2B and Figure S1B). The results indicate that modularity phenomenon of the predicted CMs resemble the modularity patterns observed in the experimental CMs. In addition, the predicted CMs successfully capture the white band regions present in the experimental CM (Figure S1B).

To establish the biological significance of COCOA model predictions, we generate plots that illustrate the predicted CMs alongside the epigenomic signal tracks and the PC1 tracks. Figure S1C reveals that the predicted CM accurately show plaid patterns of chromatin compartments, with each block of the plaid corresponding to a signal peaks in the epigenomic data tracks. The PC1 values from the tracks of the experimental CM and the predicted CM also align precisely with these results. Importantly, COCOA can infer chromatin compartments that are consistent with the underlying epigenomic data but are not captured in the experimental CM (indicated by the black dotted lines in Figure 2C). Moreover, we analysed shifts in six epigenomic modification signals at compartment boundaries and randomly selected genomic loci, as done in previous studies [54]. Notably, we observed the consistent significant shifts in epigenomic modification signals within 375 kb neighbourhoods around A/B compartment boundaries in both predicted and experimental CMs. These shifts were obviously different from randomly selected genomic loci (Figure 2D and Figure S1D). It is worth noting that shifts of partial epigenomic modification signals of the predicted CM generated by COCOA outperformed the experimental CM in capturing some compartment boundaries (e.g.: A2B boundary of H3K4me1 shown in Figure 2D).
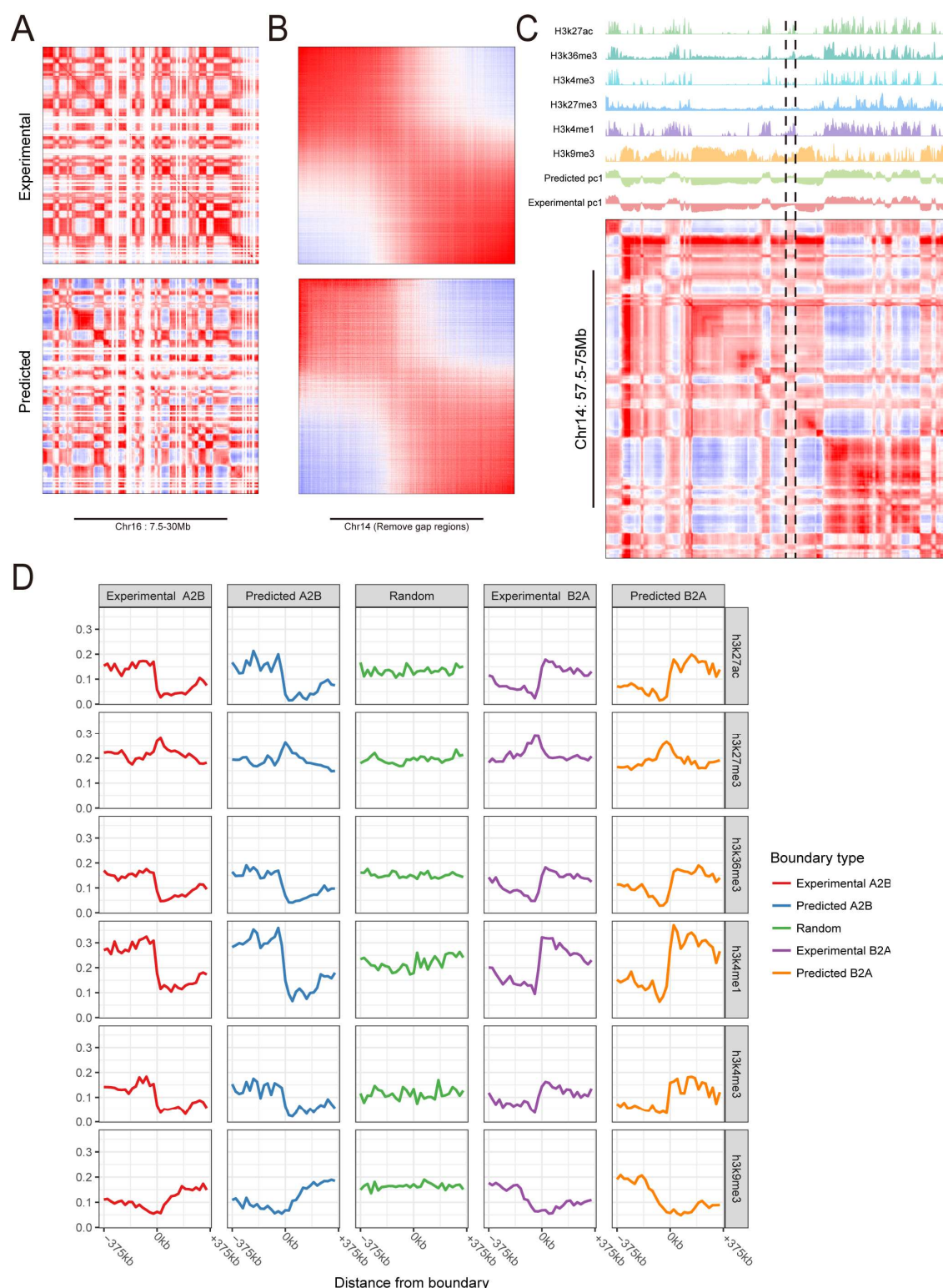
Figure 2 COCOA accurately predicts significant compartment patterns from epigenomic data

**(A)** Representative region illustrating predicted and experimental CM on test chromosomes.

(B) Heatmaps of the experimental CM and the predicted CM, sorted according to their respective PC1 sizes. The predicted CM demonstrates consistency with the compartment patterns observed in experimental CM.

(C) The predicted CM exhibits patters that align precisely with the waveform of histone modification signal. Within the region marked by the black dashed line, COCOA is able to correct the pattern misclassified by the experiment.

(**D**) Analysis of the shifts in modification signals within 375 kb neighbourhoods surrounding compartment boundaries in both predicted and experimental CMs.

Genome-wide performance evaluation of COCOA

The performance of COCOA was quantitatively analysed on genome-wide test sets. We calculated the MSE, MAE, PNSR, and SSIM scores to evaluate the robustness of error, signal-to-noise ratios and the structural similarity of COCOA on test sets. COCOA achieves competitive error and similarity scores on the test sets, exhibiting only minimal fluctuation with variations in the quality of the input data and the chromosome size. This stability indicates that COCOA performs consistently across different prediction scenarios (Figure 3C, left panel and Table S3). In addition, we adopted GenomeDISCO scores, designed to assess the reproducibility of contact maps, to validate the biological significance of the predicted CM. As shown in Figure 3C (right panel), COCOA achieved high reproducibility between the predicted CM and the experimental CM.
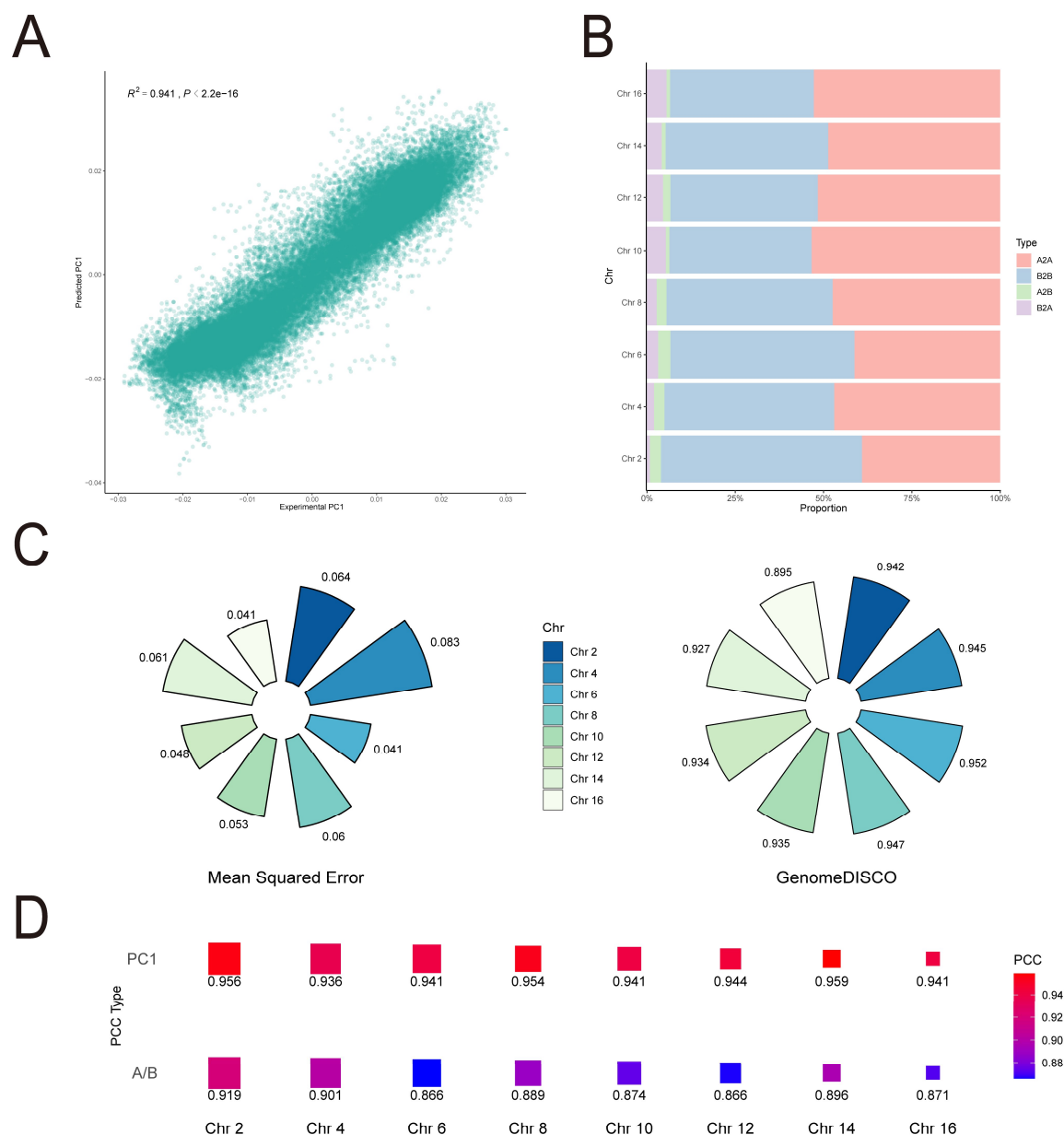
Figure 3 Performance evaluation of COCOA in multiple metrics

(A) Scatterplot showing the high correlation between the PC1 of the predicted CM and the PC1 of the ground truth on test sets.

(B) Proportion of compartment pattern matching between the predicted CM and the experimental CM. The red and blue bars represent the proportion of compartments that overlap between the predicted CM and experimental CM. The green and purple bars indicate the proportion of compartments that differ between the predicted CM and experimental CM.

(C) The MSE and genomeDISCO scores for COCOA on the test chromosomes.

(D) Correlation coefficient between the predicted CM and experimental CM. The "PC1" row represents the correlation of the first principal component of the CM and the "A/B" row represents the correlation after binarization of PC1.

As the CM contains abundant information regarding chromatin compartmentalization, we preformed correlation evaluations at the CM, PC1 values and the compartment states levels. Figure 3A shows a scatter plot of the PC1 values between the predicted CM and the experimental CM across the test sets ($R^2 = 0.941, p < 2.2e-16$). In Figure 3D, the correlation coefficients of the PC1 values for each predicted and experimental CM

are higher than 0.9, with the same trend observed for the A/B chromatin state correlation coefficients. Furthermore, we observed that misclassification rates of A/B compartments were independent of chromosome length and all remained below 10% (Figure 3B). To evaluate COCOA's performance in inferring deep chromatin compartmentalization information, we calculated the mean correlation coefficients for each column between the predicted CM and the ground truth. The results show that the predicted CM achieved high correlation scores when compared to the experimental CM (Figure S2).

COCOA predicts chromatin compartmentalization changes to epigenomic perturbations

After confirming the accuracy of COCOA in inferring chromatin compartmentalization from epigenomic data, we used COCOA to perform *in silico* epigenomic perturbation experiments and assessed the impact of epigenomic signals on chromatin compartment patterns.

In the single epigenomic marker perturbation (one-perturbation) experiments, we generated perturbed epigenomic data by setting one selected epigenomic signal to its minimum value while keeping other data unchanged. Subsequently, we predicted the corresponding CMs for the perturbed epigenomic data and compared them to the experimental CM for the unperturbed data. The predicted results from one-perturbation experiments indicated that alerting H3K9me3 signal significantly influenced chromatin architecture, causing a substantial number of compartment B-to-A switches (Figure 4A). On the other hand, the perturbation of H3K4me1 signal led to a small proportion of compartment A-to-B switches. Perturbing other epigenomic signals (e.g., H3K27ac, H3K27me3, H3K4me1, and H3K4me3) had no significant effect on chromatin compartment patterns (Figure 4A and Figure S3A).

To further analyse the contribution of individual epigenomic signals to the maintenance of chromatin compartment patterns, we conducted keep-one epigenomic signal perturbation (keep-one perturbation) experiments. In these experiments, we perturbed the epigenomic data by maintaining the selected epigenomic signal data while setting all other epigenomic signal data to their minimum values. Subsequently, we utilized COCOA to predict the chromatin compartment patterns for the perturbed data. The keep-one perturbation experiments revealed that the predicted CM from H3K9me3 or H3K4me1 signal partially overlapped with the experimental CM for the unperturbed data, while those predicted CMs from other epigenomic signals exhibited distinct differences from the experimental CM (Figure S4A and Figure S4B). This result reinforced the importance of H3K9me3 and H3K4me1 for predicting the status of chromatin architecture.

We next investigated the effects of two epigenomic signals on the chromatin compartmentalization through two epigenomic signal perturbation (two-perturbations) experiments. We observed that H3K9me3 and H3K4me1 signals play dominant roles in determining A/B compartment patterns. When H3K9me3 or H3K4me1 signal was perturbed along with H3K27ac, H3K27me3, H3K4me1 or H3K4me3 signal, the predicted chromatin compartment patterns exhibited significant changes compared to those for the unperturbed data (Figure S3B). Notably, perturbing the H3K9me3 signal gave rise to compartment B-to-A changes, while perturbing the H3K4me1 signal resulted in compartment A-to-B changes (Figure 4B). Simultaneous perturbations of H3K9me3

and H3K4me1 signal exhibited the greatest impact on the changes in chromatin compartment patterns among all the two-perturbations experiments.

Taken together, COCOA facilitates investigations into the role of epigenomic signals in determining chromatin compartmentalization through *in silico* epigenomic perturbation experiments. Our results suggest that H3K9me3 and H3K4me1 signals are crucial for maintaining the chromatin compartment patterns.
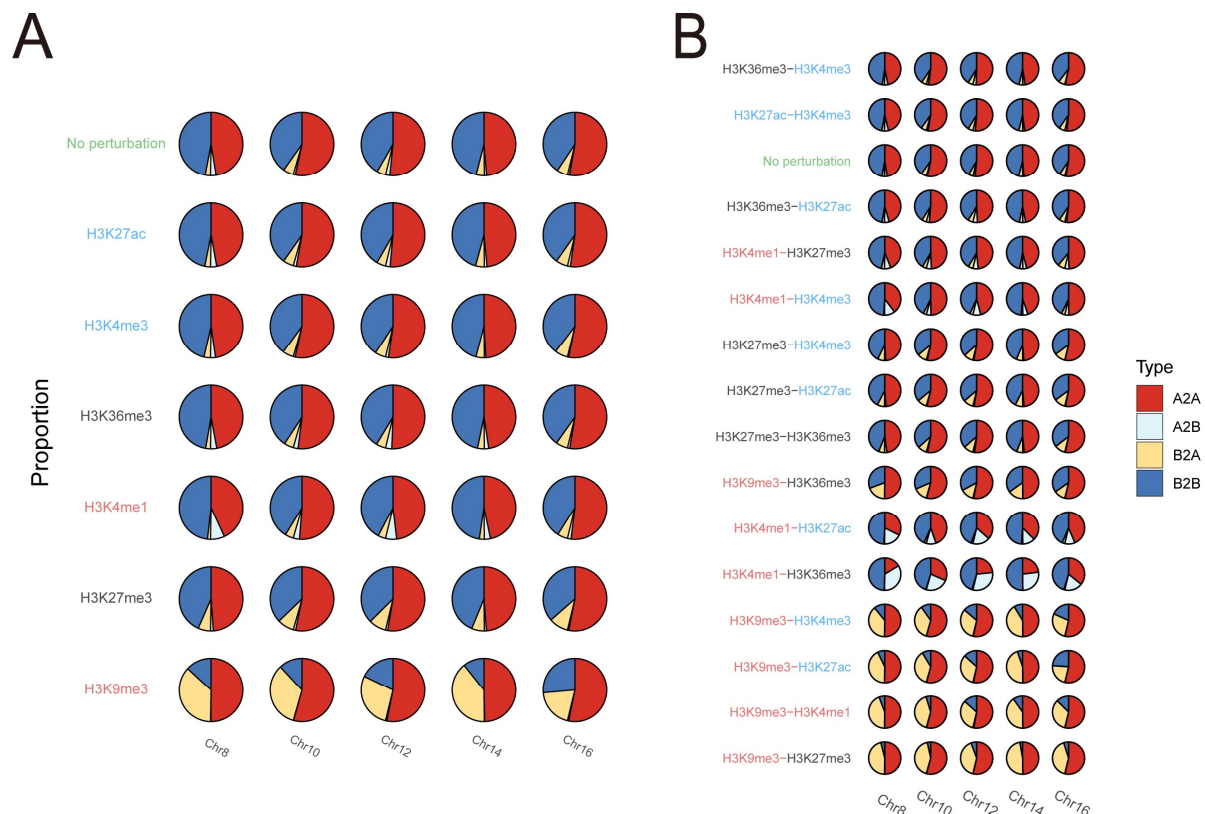
Figure 4 COCOA predicts the compartment patterns for epigenomic perturbation experiments

The row where green "No perturbation" (Reference) indicates the comparison between the predicted CM and ground truth for unperturbed data. Red, black, and blue fonts in the vertical axis labels indicate high, medium, and low impact of the perturbed epigenomic signal on the chromatin compartment patterns, respectively.

(A) Proportion of the compartment pattern matching between the predicted CM and the ground-truth CM from the 1-perturbation epigenomic combinations.

(B) Proportion of the compartment pattern matching between the predicted CM and the ground-truth CM from the 2-perturbation epigenomic combinations.

COCOA shows robust performance of model predictions at different resolutions

To evaluate the performance of COCOA at different resolutions, we used the model trained at 25kb to predict the fine-scale CM using resolution-specific inputs. We first used the trained model to predict 10kb CM for chromosome 16, 17 and 18 data sets and evaluated the performance of the model predictions. The results shown

in Table S4 indicate that COCOA achieves consistent and competitive scores on all three test sets. We further evaluated the correlations between the predicted compartment patterns and the ground truth patterns. The predicted 10 kb CM is highly correlated with and the experimental CM (Figure 5A). A similar high correlation was observed between the predicted compartment states and the ground truth states (PC1 of CM). The compartment misclassification rate indicates that COCOA generates CMs containing reliable chromatin compartment information (Figure S5A). The predicted CM exhibits similar plaid patterns to the experimental CM, which correspond well with the epigenomic signals (Figure 5B and Figure 5C). In addition, the modularity of the predicted CM aligns with that of the experimental CM on the whole (Figure S5C). We also observed consistent and significant shifts in epigenomic modification signals within 150kb neighbourhoods around A/B compartment boundaries, in both predicted and experimental CM, which are distinguishable from randomly selected genomic loci (Figure 5D).
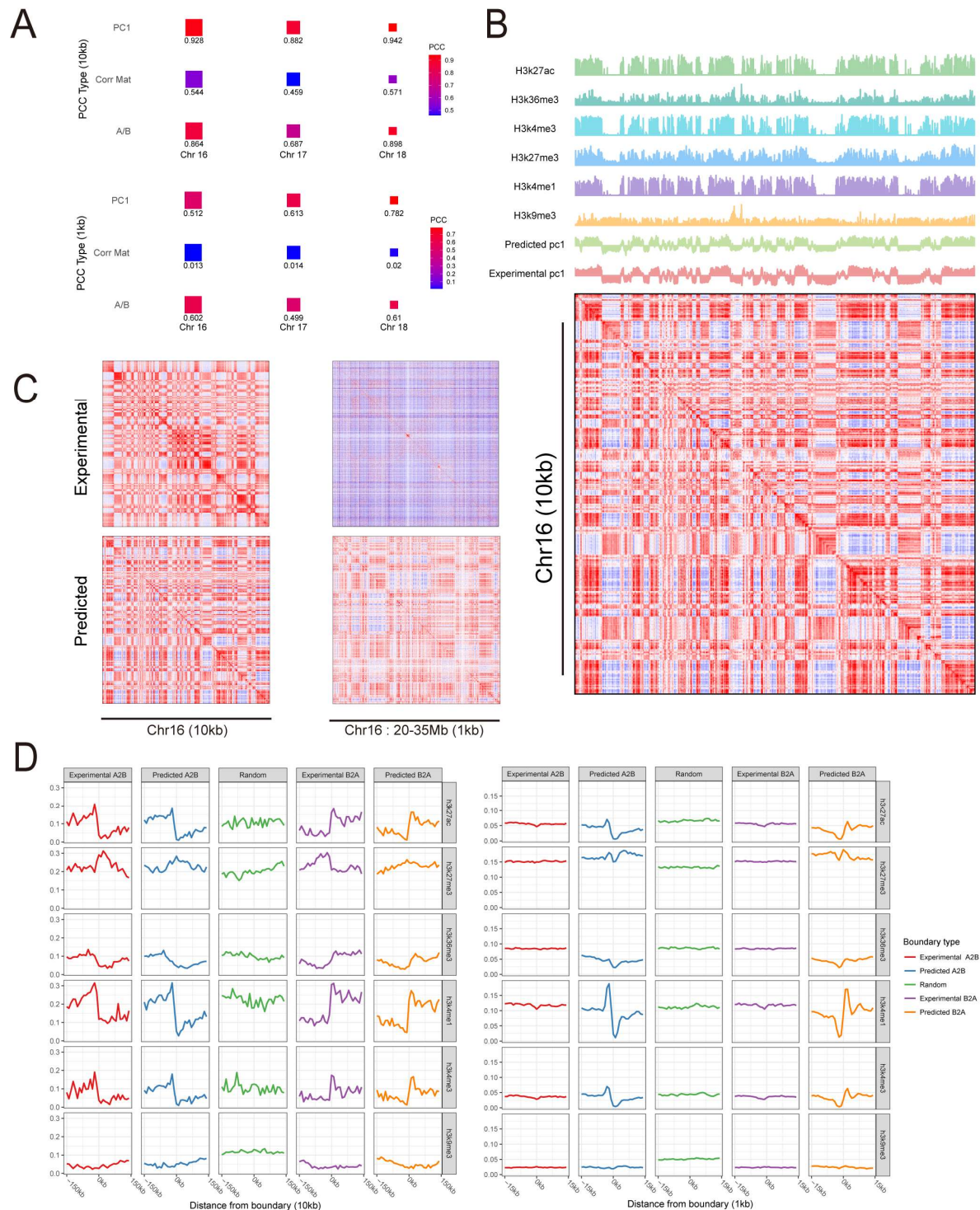
Figure 5 Predicting resolution-specific compartment pattern with COCOA

(A) Correlation coefficient between the predicted CM and the experimental CM at 10kb and 1kb resolutions. The rows labeled 'PC1', 'Corr Mat' and 'A/B' represent the correlation of the PC1 of the CMs, the correlation between the predicted CM and the experimental CM, and the correlation of the binarized PC1s, respectively.

(B) The predicted CM accurately capture the histone modification signal waveforms for chr16 at 10kb resolution.

(C) Typical region of the predicted CM and the experimental CM at 10kb and 1kb resolutions. At 1kb resolution, the experimental CM exhibits high noise levels and lacks a recognizable plaid pattern, while the predicted CM demonstrates a clear plaid pattern

(D) Shifts observed in modification signals around compartment boundaries in both predicted and experimental CMs at 10 kb and 1 kb resolutions. At 10 kb resolution, both predicted and experimental CMs display meaningful shifts, whereas at 1 kb resolution, the experimental CM approaches random results, while the predicted CM still shows significant biological shifts.

To evaluate COCOA's performance at ultra-high resolution, we employ it to predict 1kb for chromosome 16, 17 and 18, using the model trained at 25kb resolution. Similar to the evaluation at 10kb resolution, we assessed the performance metrics and correlations. The results show that COCOA achieves robust performance across a wide range of scores, but obtains scores close to 0 for the CM correlation coefficient (Figure 5A, Table S4). This may be attributed to the sparsity of the deeply-sequenced experimental CM at ultra-high resolution (i.e., ~2.6–4.5 billion uniquely mapped reads with ~150X coverage per nucleosome) [47], which is challenging to define as the ground truth. As the CM size increases, the mean error evaluation narrows the gap, producing similar scores. However, using PCA-based correlation or compartment misclassification rates, we can partially mitigate the sparsity issue and obtain reliable scores (Figure S5B). Therefore, we visualized the experimental and predicted CM by heatmaps (Figure 5C). We found that the experimental CM shows vaguely visible plaid patterns and is filled with noise-induced thin lines. In contrast, the predicted CM remains consistent with these fuzzy patterns but displays more apparent compartmentalization patterns (Figure 5C and Figure S5D). Moreover, we investigated the shifts of epigenomic modification signals between the predicted CM and the experimental CM. Surprisingly, the shifts observed in the experimental CM are similar to those in randomly selected genome loci, while the predicted CM shows significant and biologically meaningful shifts (Figure 5D).

COCOA accurately predicts cell-type-specific chromatin compartment patterns

Because epigenomic data are cell-type-specific, we tested whether COCOA can accurately predict chromatin compartment patterns in different cell types. We first applied COCOA to GM12878 datasets and generated the predicted CM for multiple chromosomes. The experimental CM obtained from the Hi-C data of GM12878 [6] served as the ground truth for comparison. The results indicated that high correlations were observed between the two CMs and between the two principal components (PC1s) of the CMs (Figure 6A and Figure S6A). Figure 6B revealed that the compartment misclassification rates were all below 20%. Furthermore, the predicted CM presents the plaid patterns consistent with the experimental CM, achieving stable and competitive scores in terms of both error and the image similarity (Figure 6C and Table S5). Taken together, these data suggest that COCOA reliably predicts the cell-type-specific chromatin compartment patterns.

Having established COCOA's capability to predict CMs across diverse cell types, we proceeded to predict CMs in five additional cell types encompassing tissues, diseases, and primary cells where chromatin conformation had not been sequenced. The predicted CM for the five data sets displayed obvious plaid patterns (Figure 6D and Figure S6B). The predicted A/B compartments in SJSA1, NCI-H929, and activated B cells were clearly defined and noise-free, while the predicted CMs of activated T cells and chorionic villus tissues displayed slightly diminished performance. To gain insight into the effects of chromatin compartmentalization in disease and differentiation, we systematically compared the patterns across different cell types. Using predicted CM from GM12878 data as a benchmark, we examined commonalities and differences in chromatin region patterns based on histone modification track information (Figure 6E). The results revealed similar patterns between GM12878

and most other cells, albeit with variations in certain regions (Figure 6F). Notably, activated B cells, being immune cells akin to GM12878, exhibited a comparable compartment pattern. Similarly, activated T cells demonstrated an analogous pattern. In contrast, SJSA1, NCI-H929, and chorionic villus tissues exhibited distinct compartment patterns.
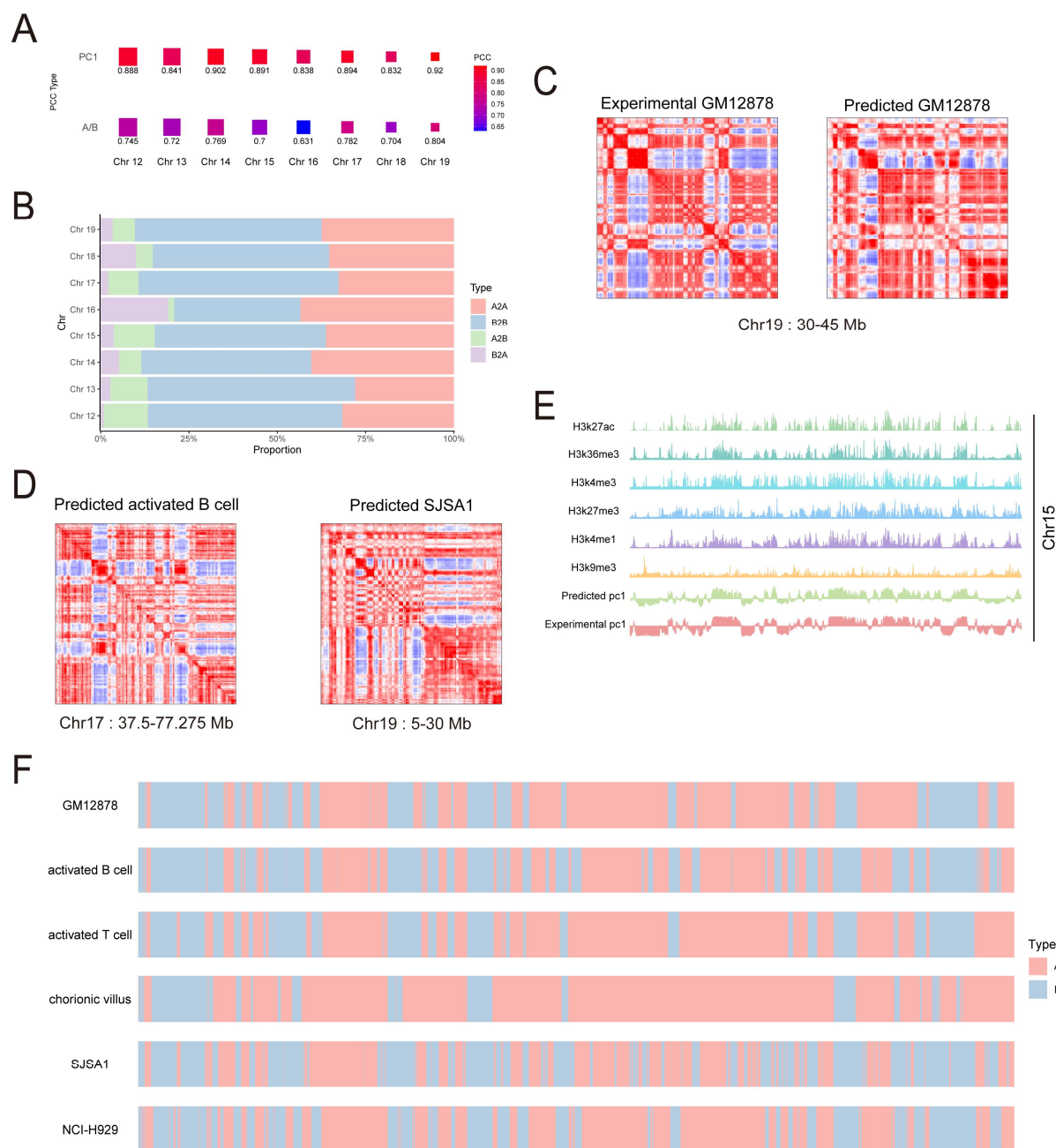
Figure 6 Cell-type-specific prediction of compartment pattern by COCOA

(A) Correlation coefficient between the predicted CM and the experimental CM on GM12878 datasets. The rows 'PC1' and 'A/B' represent the correlation of CM's PC1s and the correlation of binarized PC1s, respectively.

(B) Proportion of matching compartment patterns between the predicted CM and the experimental CM on GM12878 datasets.

(C) Typical regions showing the predicted CM and the experimental CM patterns on GM12878 datasets.

(D) Example regions illustrating the predicted CM patterns on SJSA1 and activated B cells datasets.

(E) Precisely matching of predicted CM patterns with the waveform of histone modification signals on chr15 in GM12878 datasets.

(F) Systematic comparison of chromatin compartment status on chr15 across GM12878, activated B cells, activated T cells, chorionic villus, SJSA1 and NCI-H929 datasets.

Discussion

In this study, we developed a deep neural network framework called COCOA, which incorporates six types of epigenomic modification signals to accurately predict fine-scale resolution chromatin compartment patterns. These epigenomic signals data are readily accessible in ENCODE database [53] for various cell lines, in vitro differentiated cells, primary cells, and tissues. To process the raw epigenomic data, we employed resolution-specific pre-processing to bin the data into mated inputs from different genomic positions. COCOA then uses the bidirectional feature reconstruction module to extract track features from these mated inputs, then fuses these track features to contact features using the cross attention fusion module. Eventually, these contact features are converted to chromatin compartment patterns by the residual feature reduction module. COCOA predict directly long-range chromatin compartment patterns without considering short-range interactions [28,29,35]. Our results demonstrate that COCOA accurately predicts the same chromatin compartment patterns as the experimental CM, with consistent epigenomic signals shifts of these patterns (Figure 2D). During model evaluation, COCOA achieves excellent performance with robust reproducibility scores on the test sets. Furthermore, the predicted CM and its PC1 show a high correlation with the experimental CM and PC1. The compartment misclassification rates of the predicted CM remain below 10% and are independent of chromosome length.

With COCOA's accurate prediction of chromatin compartmentalization, it becomes possible to perform *in silico* epigenomic perturbation to study the influence of histone modification signals on chromatin compartmentalization. By generating predicted CM using different perturbed epigenomic data, we found that H3K9me3 has strong impact on chromatin compartment patterns, followed by H3K4me1. In contrast, H3K27me3 and H3K36me3 have a moderate level of impact, and H3K27ac and H3K4me3 have low impact. Interestingly, COCOA predicted that perturbation of H3K9me3 signal led to compartment B-to-A changes, while perturbation of H3K4me1 signal resulted in compartment A-to-B switches. Additionally, H3K9me3 and H3K4me1signals play dominant roles in determining chromatin compartment patterns when they are perturbed together with other epigenomics signals in two-perturbations experiments.

Furthermore, we explored the performance of COCOA's predictions across different resolutions and cell types. For prediction at 10kb resolution, COCOA exhibited the same outstanding performance as predicted at the trained resolution (25kb). Recognizing the significance of high resolution in chromatin interaction data analysis, we investigated whether COCOA can make good prediction at 1kb resolution. Unfortunately, even with a high sequenced depth [47], the experimental CM at 1kb resolution contains excessive noise lines and barely discernible plaid pattern. Therefore, we analysed histone modification shifts at compartment boundaries and mapped heatmaps of the predicted CM at different genome ranges. Surprisingly, the predicted CM displayed clearer plaid patterns and exhibited more biologically meaningful shifts compared to the experimental CM and randomly selected loci. We then evaluated the performance of COCOA in predicting cell-type-specific compartment patterns. Using validated Hi-C data of GM12878, our results demonstrated that COCOA can correctly infer chromatin compartment patterns from epigenomic data on unseen cell lines.

While this work presents promising results, it also has several potential areas for improvements. Firstly, as a data-driven approach, COCOA relies on moderately good quality training sets to achieve high performance by incorporating potential information from bidirectional epigenomic data. In addition, we observed that the transfer capacity of COCOA in cross cell lines experiments is affected by the epigenomic data quality. Developing new data processing schemes may prove beneficial in solving this issue. Secondly, in challenging task such as high-

volume fine-scale resolution CM prediction and *in silico* epigenomic signal perturbation experiments, COCOA requires significant run time and substantial computational resources. To alleviate this computational burden, parallel CM generation and distributed implementations can be explored as feasible approaches [55]. Thirdly, we also preliminarily explored the influences of histone modification signals on A/B chromatin compartmentalization in HFF datasets by *in silico* epigenomic perturbation experiments. However, more systematically studying the combined impacts of epigenomic modifications in relation to complex chromatin compartmentalization on different cell lines would benefit from further experimental evidence. Lastly, COCOA's predictions for fine-scale chromatin compartmentalization information in diseases, tissues and primary cells have not been thoroughly explored. In future, it would be interesting to explore the impact of the chromatin compartment alteration on cell differentiation and disease occurrence by integrating epigenomics data with other omics and phenotypic data.

Code availability

The source codes were implemented in Python and can be freely accessed on GitHub at the following link:

https://github.com/onlybugs/COCOA

Data availability

The datasets used in this study are listed in Table S1, S2 and can be obtained from the ENCODE project and the 4D Nucleome Data Portal.

CRediT author statement

Kai Li: Conceptualization, Methodology, Software, Data curation, Visualization，Formal analysis, Writing - original draft. **Ping Zhang:** Conceptualization, Writing - review. **Jinsheng Xu:** Data curation. **Zi Wen:** Data curation. **Junying Zhang:** Data curation. **Zhike Zi:** Writing - review & editing, Resources, Conceptualization. **Li Li:** Writing - review & editing, Project administration, Resources. All authors have read and approved the final manuscript.

Competing interests

The authors have declared no competing interests.

Acknowledgments

ORCID

ORCID 0000-0002-9594-8341 (Li Li)

ORCID 0000-0002-7601-915X (Zhike Zi)

ORCID 0000-0001-6831-1807 (Ping Zhang)

References

[1]   Ke Y, Xu Y, Chen X, Feng S, Liu Z, Sun Y, et al. 3D Chromatin Structures of Mature Gametes and Structural Reprogramming during Mammalian Embryogenesis. Cell 2017;170:367-381.e20.

[2]   Li T, Li R, Dong X, Shi L, Lin M, Peng T, et al. Integrative Analysis of Genome, 3D Genome, and Transcriptome Alterations of Clinical Lung Cancer Samples. Genomics Proteomics Bioinformatics 2021;19:741–53.

[3]   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.

[4]   Hsieh T-HS, 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.

[5]   Deshpande AS, Ulahannan N, Pendleton M, Dai X, Ly L, Behr JM, et al. Identifying synergistic high-order 3D chromatin conformations from genome-scale nanopore concatemer sequencing. Nat Biotechnol 2022.

[6]   Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 2014;159:1665–80.

[7]   Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature 2012;485:381–5.

[8]   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.

[9]   Vian L, Pękowska A, Rao SSP, Kieffer-Kwon K-R, Jung S, Baranello L, et al. The Energetics and Physiological Impact of Cohesin Extrusion. Cell 2018;175:292–4.

[10] Goel VY, Huseyin MK, Hansen AS. Region Capture Micro-C reveals coalescence of enhancers and promoters into nested microcompartments. Nat Genet 2023;55:1048–56.

[11] Simonis M, Klous P, Splinter E, Moshkin Y, Willemsen R, de Wit E, et al. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). Nat Genet 2006;38:1348–54.

[12] Spracklin G, Abdennur N, Imakaev M, Chowdhury N, Pradhan S, Mirny LA, et al. Diverse silent chromatin states modulate genome compartmentalization and loop extrusion barriers. Nat Struct Mol Biol 2023;30:38–51.

[13] Haws SA, Simandi Z, Barnett RJ, Phillips-Cremins JE. 3D genome, on repeat: Higher-order folding principles of the heterochromatinized repetitive genome. Cell 2022;185:2690–707.

[14] Feng Y, Wang Y, Wang X, He X, Yang C, Naseri A, et al. Simultaneous epigenetic perturbation and genome imaging reveal distinct roles of H3K9me3 in chromatin architecture and transcription. Genome Biol 2020;21:296.

[15] Nichols MH, Corces VG. Principles of 3D compartmentalization of the human genome. Cell Rep 2021;35:109330.

[16] Wen Z, Zhang W, Zhong Q, Xu J, Hou C, Qin ZS, et al. Extensive Chromatin Structure-Function Associations Revealed by Accurate 3D Compartmentalization Characterization. Front Cell Dev Biol 2022;10:845118.

[17] Harris HL, Gu H, Olshansky M, Wang A, Farabella I, Eliaz Y, et al. Chromatin alternates between A and B compartments at kilobase scale for subgenic organization. Nat Commun 2023;14:3303.

[18] Rao SSP, Huang S-C, Glenn St Hilaire B, Engreitz JM, Perez EM, Kieffer-Kwon K-R, et al. Cohesin Loss Eliminates All Loop Domains. Cell 2017;171:305-320.e24.

[19] Fortin J-P, Hansen KD. Reconstructing A/B compartments as revealed by Hi-C using long-range correlations in epigenetic data. Genome Biol 2015;16:180.

[20] Schmitt AD, Hu M, Ren B. Genome-wide mapping and analysis of chromosome architecture. Nat Rev Mol Cell Biol 2016;17:743–55.

[21] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436–44.

[22] Yang JY, Chang J-M. Pattern recognition of topologically associating domains using deep learning. BMC Bioinformatics 2022;22:634.

[23] Soler-Vila P, Cuscó P, Farabella I, Di Stefano M, Marti-Renom MA. Hierarchical chromatin organization detected by TADpole. Nucleic Acids Res 2020;48:e39–e39.

[24] Zhang S, Plummer D, Lu L, Cui J, Xu W, Wang M, et al. DeepLoop robustly maps chromatin interactions from sparse allele-resolved or single-cell Hi-C data at kilobase resolution. Nat Genet 2022;54:1013–25.

[25] Zhang Y, Blanchette M. Reference panel guided topological structure annotation of Hi-C data. Nat Commun 2022;13:7426.

[26] Wang B, Liu K, Li Y, Wang J. DFHiC: a dilated full convolution model to enhance the resolution of Hi-C data. Bioinformatics 2023;39:btad211.

[27] Zhang Y, An L, Xu J, Zhang B, Zheng WJ, Hu M, et al. Enhancing Hi-C data resolution with deep convolutional neural network HiCPlus. Nat Commun 2018;9:750.

[28] Schwessinger R, Gosden M, Downes D, Brown RC, Oudelaar AM, Telenius J, et al. DeepC: predicting 3D genome folding using megabase-scale transfer learning. Nat Methods 2020;17:1118–24.

[29] Fudenberg G, Kelley DR, Pollard KS. Predicting 3D genome folding from DNA sequence with Akita. Nat Methods 2020;17:1111–7.

[30] Yang R, Das A, Gao VR, Karbalayghareh A, Noble WS, Bilmes JA, et al. Epiphany: predicting Hi-C contact maps from 1D epigenomic signals. Genome Biol 2023;24:134.

[31] Zhang R, Zhou T, Ma J. Multiscale and integrative single-cell Hi-C analysis with Higashi. Nat Biotechnol 2022;40:254–61.

[32] Xiong K, Zhang R, Ma J. scGHOST: Identifying single-cell 3D genome subcompartments. Bioinformatics; 2023.

[33] Zhang S, Chasman D, Knaack S, Roy S. In silico prediction of high-resolution Hi-C interaction matrices. Nat Commun 2019;10:5449.

[34] Zhou J. Sequence-based modeling of three-dimensional genome architecture from kilobase to chromosome scale. Nat Genet 2022;54:725–34.

[35] Tan J, Shenker-Tauris N, Rodriguez-Hernaez J, Wang E, Sakellaropoulos T, Boccalatte F, et al. Cell-type-specific prediction of 3D chromatin organization enables high-throughput in silico genetic screening. Nat Biotechnol 2023.

[36] Reiff SB, Schroeder AJ, Kırlı K, Cosolo A, Bakker C, Mercado L, et al. The 4D Nucleome Data Portal as a resource for searching and visualizing curated nucleomics data. Nat Commun 2022;13:2365.

[37] Abdennur N, Mirny LA. Cooler: scalable storage for Hi-C data and other genomically labeled arrays. Bioinforma Oxf Engl 2020;36:311–6.

[38] Kim TH, Dekker J. ChIP-seq. Cold Spring Harb Protoc 2018.

[39] Colwell J. Expanding the Scope of ENCODE. Cancer Discov 2016;6:OF4.

[40] Li K, Zhang P, Wang Z, Shen W, Sun W, Xu J, et al. iEnhance: a multi-scale spatial projection encoding network for enhancing chromatin interaction data resolution. Brief Bioinform 2023;24:bbad245.

[41] Wang X, Qin Y, Wang Y, Xiang S, Chen H. ReLTanh: An activation function with vanishing gradient resistance for SAE-based DNNs and its application to rotating machinery fault diagnosis. Neurocomputing 2019;363:88–98.

[42] Dai Y, Gieseke F, Oehmcke S, Wu Y, Barnard K. Attentional Feature Fusion 2020.

[43] He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition 2015.

[44] Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift 2015.

[45] Johnson J, Alahi A, Fei-Fei L. Perceptual Losses for Real-Time Style Transfer and Super-Resolution 2016.

[46] Gatys LA, Ecker AS, Bethge M. A Neural Algorithm of Artistic Style 2015.

[47] Krietenstein N, Abraham S, Venev SV, Abdennur N, Gibcus J, Hsieh T-HS, et al. Ultrastructural Details of Mammalian Chromosome Architecture. Mol Cell 2020;78:554-565.e7.

[48] Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library 2019.

[49] Kingma DP, Ba J. Adam: A Method for Stochastic Optimization 2017.

[50] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python 2018.

[51] Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process Publ IEEE Signal Process Soc 2004;13:600–12.

[52] Yang T, He X, An L, Li Q. Methods to Assess the Reproducibility and Similarity of Hi-C Data. Methods Mol Biol Clifton NJ 2022;2301:17–37.

[53] Luo Y, Hitz BC, Gabdank I, Hilton JA, Kagda MS, Lam B, et al. New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. Nucleic Acids Res 2020;48:D882–9.

[54] Xiong K, Ma J. Revealing Hi-C subcompartments by imputing inter-chromosomal chromatin interactions. Nat Commun 2019;10:5069.

[55] Hu Y, Ma W. EnHiC: learning fine-resolution Hi-C contact maps using a generative adversarial framework. Bioinformatics 2021;37:i272–9.