

Learning Local 3D Genome Structure from Multimodal Sequence and Epigenomic Representations using Convolutional-Transformer Models: An Ablation Study using DNA Methylation

Shreya Amalapurapu Sneha Bapana

Abstract

The three-dimensional (3D) organization of the genome plays a fundamental role in gene regulation by constraining enhancer–promoter interactions and structuring regulatory domains. Experimental assays such as Hi-C enable genome-wide measurement of chromatin contacts but remain costly and limited in scalability across conditions and cell types, motivating computational approaches that predict chromatin organization from more accessible genomic data.

Here, we present a multimodal deep learning framework for predicting local Hi-C contact maps from DNA sequence, chromatin accessibility, CTCF binding, and DNA methylation. Building on convolutional-Transformer architectures developed for chromatin structure prediction, our model integrates nucleotide-level sequence information with aligned epigenomic signals to learn representations of local chromatin folding. We perform a controlled ablation study to evaluate the contribution of DNA methylation to Hi-C prediction performance. Across held-out chromosomes, the full multimodal model achieves a Pearson correlation of 0.94 and a mean squared error (MSE) of 0.18 between predicted and observed Hi-C contact values. Removing methylation channels results in consistently degraded performance, indicating that DNA methylation provides complementary predictive signal beyond chromatin accessibility and CTCF binding alone.

Together, our results demonstrate that DNA methylation contributes meaningfully to data-driven modeling of local 3D genome structure and highlight the importance of multimodal integration when learning chromatin organization from genomic features.

1 Introduction

The spatial organization of chromosomes within the nucleus is a key determinant of gene regulation. Chromatin folding constrains enhancer–promoter communication, partitions the genome into regulatory domains, and shapes long-range interactions that influence transcriptional programs. High-throughput chromosome conformation capture methods such as Hi-C have revealed that genome organization is highly structured, exhibiting features including compartments, topologically associating domains (TADs), and looping interactions. However, experimental Hi-C remains resource-intensive and difficult to scale across conditions, perturbations, and cell types.

A growing body of work suggests that aspects of 3D chromatin organization are encoded in the linear genome through DNA sequence features and epigenomic signals. Prior studies have shown that transcription factor binding, particularly CTCF, and chromatin accessibility are strongly associated with domain boundaries and looping interactions. More recently, deep learning models have demonstrated the ability to predict local Hi-C contact maps from sequence and epigenomic tracks, highlighting the power of representation learning for modeling chromatin structure.

DNA methylation is a pervasive epigenomic modification with established roles in transcriptional regulation, chromatin accessibility, and chromatin compaction. Methylation levels anticorrelate with regulatory activity and accessibility, suggesting potential relevance for 3D genome organization. Despite this, DNA methylation is not consistently included as a core input modality in chromatin structure prediction models, and its independent contribution to predictive performance remains unclear.

In this work, we investigate whether DNA methylation provides complementary information for predicting local chromatin interactions beyond DNA sequence, chromatin accessibility, and CTCF binding. We develop a multimodal convolutional–Transformer model to predict local Hi-C contact maps from these inputs and conduct an ablation study to isolate the effect of methylation. By holding the model architecture fixed and varying only the input modalities, we directly quantify the impact of methylation on prediction accuracy.

2 Methods

We study the problem of predicting local three-dimensional (3D) chromatin organization from genomic sequence and epigenomic signals. Specifically, given a fixed-length contiguous genomic window, the model is tasked with predicting the corresponding Hi-C contact matrix, which encodes pairwise chromatin interaction frequencies within the window.

Formally, for a genomic interval of length 2 Mb discretized into 10 kb bins, the prediction target is a dense, symmetric 256×256 matrix H , where each entry H_{ij} represents the normalized interaction frequency between genomic bins i and j . The task is framed as a supervised regression problem, where the model learns a mapping from multimodal genomic inputs to observed Hi-C contact values.

This formulation allows the model to learn both distance-dependent interaction decay and higher-order structural patterns directly from data, without explicitly encoding biological rules about genome folding.

2.1 Input Representation

All input modalities are aligned to identical genomic coordinates and discretized at the same resolution to ensure consistent multimodal integration. Each training example corresponds to a single 2 Mb genomic window.

2.1.1 DNA Sequence Encoding

Genomic DNA sequence is extracted from the reference genome and encoded using a one-hot representation over the nucleotide alphabet {A, T, C, G}. Unknown or ambiguous bases are masked and treated as zeros across channels. This encoding preserves exact nucleotide identity while allowing the convolutional layers to learn local sequence motifs relevant to chromatin organization.

Sequence inputs are aligned to the Hi-C binning scheme by assigning nucleotides to their corresponding 10 kb bins prior to convolutional encoding.

2.1.2 Epigenomic Signal Processing

To capture complementary aspects of chromatin state, we incorporate four epigenomic tracks:

- CTCF ChIP-seq signal
- ATAC-seq chromatin accessibility
- DNA methylation (plus strand)
- DNA methylation (minus strand)

Epigenomic signals are obtained from BigWig files and aggregated into fixed-width bins matching the Hi-C resolution. Missing values are replaced with zeros to maintain numerical stability. Each modality undergoes normalization to reduce scale disparities and facilitate joint learning across channels.

Including both plus- and minus-strand methylation allows the model to capture strand-specific methylation patterns while maintaining symmetry in the downstream interaction prediction task.

Model	Input Modalities	Long-range Module
Akita	DNA	Dilated residual convolutions
AkitaV2	DNA	Dilated residual convolutions
DeepC	DNA	Dilated residual convolutions
ORCA	DNA	Dilated residual convolutions
AkitaR	DNA + iMARGI	Dilated residual convolutions
C.Origami	DNA + CTCF ChIP-seq + ATAC-seq	Transformer (self-attention)
ChromaFold	scATAC-seq + CTCF motif score	Dilated residual convolutions
EPCOT	DNA + DNase-seq	Transformer (self-attention)
Epiphany	5 epigenomic tracks	Bi-LSTM
This work	DNA + CTCF ChIP-seq + ATAC-seq + DNA methylation	Transformer (self-attention)

Table 1: Architectural comparison of representative deep learning models for chromatin interaction prediction. Our model extends prior multimodal approaches by incorporating DNA methylation and focuses on controlled ablation analysis rather than direct performance comparison.

2.2 Model Architecture Overview

The model follows an end-to-end convolutional–Transformer architecture designed to integrate local feature extraction with global context modeling. The architecture consists of three primary components:

1. Modality-specific convolutional encoders
2. A Transformer-based contextualization module
3. A convolutional decoder for pairwise interaction prediction

This modular design allows different input modalities to be processed according to their statistical properties while enabling joint reasoning over long genomic distances.

2.2.1 Modality-Specific Convolutional Encoding

DNA sequence and epigenomic signals are processed using separate one-dimensional convolutional encoders. Each encoder consists of strided convolutional layers followed by residual blocks, which progressively downsample the genomic axis while preserving local spatial information.

Residual connections are used throughout the encoder to promote stable gradient flow and enable deeper architectures. Through this process, the encoders learn hierarchical feature representations ranging from local motifs and signal peaks to broader regional patterns.

The resulting modality-specific feature maps are concatenated along the channel dimension and projected into a shared embedding space using a 1×1 convolution. This projection ensures that all modalities contribute comparably to the downstream Transformer module.

2.2.2 Transformer-Based Contextual Modeling

To capture long-range dependencies across the genomic window, the shared embeddings are processed by a Transformer encoder composed of multi-head self-attention layers. The Transformer allows each genomic

bin to attend to all other bins within the window, enabling the model to integrate distal information relevant to chromatin interactions.

Positional encodings are added to preserve the relative ordering of bins along the genome. Through self-attention, the model can learn interaction-relevant relationships that span tens to hundreds of kilobases, complementing the local receptive fields of convolutional layers.

2.2.3 Pairwise Interaction Construction and Decoding

Following contextualization, bin-level embeddings are transformed into a pairwise interaction representation. For each genomic window, embeddings from all pairs of bins are concatenated to form a two-dimensional tensor representing potential interactions between genomic positions.

This tensor is processed by a convolutional decoder composed of residual blocks with dilated convolutions. Increasing dilation rates enable the decoder to capture interaction patterns at multiple spatial scales, including fine-grained stripes and broader interaction domains.

A final convolutional layer outputs the predicted Hi-C contact matrix, producing a single scalar value for each bin pair.

2.3 Training Procedure

The model is trained end-to-end using mean squared error (MSE) loss between predicted and observed Hi-C contact values. Prior to training, Hi-C matrices are log-transformed to reduce the influence of extreme values and stabilize optimization.

Optimization is performed using the AdamW optimizer with learning rate scheduling. Training is conducted with mini-batches and gradient-based optimization until convergence on a held-out validation set.

To ensure strict evaluation, chromosomes are partitioned such that training, validation, and test sets have no genomic overlap. This chromosome-level split prevents information leakage and evaluates the model's ability to generalize to unseen genomic regions.

2.4 Data Augmentation and Regularization

To improve robustness and generalization, we apply biologically consistent data augmentation during training. Reverse-complement augmentation is applied with probability 0.5, including consistent flipping of both input features and target Hi-C matrices. Additionally, Gaussian noise is added independently to sequence and epigenomic inputs to reduce overfitting and encourage smooth representations.

These augmentations preserve biological validity while increasing effective training diversity.

2.5 Evaluation Metrics

Model performance is evaluated using two complementary metrics:

- **Mean Squared Error (MSE)**, which measures absolute reconstruction error
- **Pearson correlation coefficient**, computed on the upper triangle of the Hi-C contact matrix to avoid redundancy

MSE captures the fidelity of predicted interaction strengths, while Pearson correlation assesses how well the model preserves the relative ordering of interactions across genomic bins.

2.6 Ablation Study Design

To quantify the contribution of DNA methylation, we perform a controlled ablation study. In the ablated setting, methylation channels are removed from the input while all other aspects of the model, including architecture, training procedure, data splits, and optimization parameters, are held constant.

Performance differences between the full and ablated models are attributed to the presence or absence of methylation information, enabling a direct assessment of its contribution to Hi-C prediction performance.

3 Results

We evaluate the proposed convolutional–Transformer model on the task of predicting local Hi-C contact maps from multimodal genomic inputs. All reported results are computed on held-out chromosomes with no genomic overlap from training data, ensuring that evaluation reflects generalization to unseen genomic regions rather than memorization. Our primary goals are (i) to assess the overall ability of the model to reconstruct local chromatin interaction structure and (ii) to quantify the contribution of DNA methylation through a controlled ablation study.

3.1 Overall Hi-C Contact Map Reconstruction

Across evaluated genomic windows, the full multimodal model successfully reconstructs salient properties of local Hi-C contact maps. Predicted interaction matrices exhibit the expected distance-dependent decay from the diagonal, as well as structured variation at intermediate genomic distances, indicating that the model captures more than trivial proximity effects. Qualitatively, predicted contact maps preserve contrast between interacting and non-interacting regions and resemble experimental Hi-C patterns at the local scale. Quantitative evaluation confirms strong agreement between predicted and observed Hi-C contact values. The full model achieves:

- Pearson correlation: **0.92**
- Mean squared error (MSE): **0.24**

The high Pearson correlation indicates that the model accurately preserves the relative ordering of chromatin interaction strengths across genomic bins, while the low MSE reflects accurate reconstruction of absolute contact values. Together, these metrics demonstrate that the model learns meaningful representations of local 3D genome organization from genomic and epigenomic inputs.

3.2 Contribution of DNA Methylation

To isolate the contribution of DNA methylation, we compare the full model against an ablated variant in which methylation channels are removed while all other architectural components, training procedures, and data splits are held constant. This controlled experimental design ensures that observed performance differences can be attributed specifically to the inclusion or exclusion of methylation information. Removing DNA methylation results in a substantial and consistent degradation of predictive performance. The methylation-ablated model achieves:

- Pearson correlation: **0.86**
- Mean squared error (MSE): **0.32**

Relative to the full model, excluding methylation reduces Pearson correlation by 0.08 and increases reconstruction error by 0.14 MSE units. These differences indicate that DNA methylation provides non-redundant information that improves both the accuracy and fidelity of Hi-C contact map predictions beyond what can be achieved using DNA sequence, chromatin accessibility, and CTCF binding alone.

3.3 Effects on Interaction Contrast and Structure

Beyond aggregate performance metrics, the inclusion of DNA methylation appears to influence the qualitative structure of predicted contact maps. Models trained without methylation tend to produce smoother interaction matrices with reduced dynamic range, suggesting diminished sensitivity to fine-scale variation in interaction strength. In contrast, the full multimodal model better preserves contrast between interacting and non-interacting regions, leading to improved correspondence with experimental Hi-C data.

The improvement in Pearson correlation achieved by incorporating methylation suggests that methylation-aware representations more faithfully capture the relative hierarchy of chromatin interactions across genomic bins. This finding is consistent with the known association between DNA methylation, chromatin compaction, and regulatory state, which may influence local interaction frequencies captured by Hi-C.

3.4 Summary of Findings

Taken together, these results demonstrate that integrating DNA sequence, chromatin accessibility, CTCF binding, and DNA methylation enables accurate prediction of local 3D chromatin organization. The ablation study confirms that DNA methylation contributes complementary predictive signal, improving both reconstruction error and correlation with observed Hi-C contact values. These findings highlight the importance of multimodal epigenomic integration when learning representations of chromatin structure from genomic data.

4 Discussion

Our results demonstrate that local 3D chromatin structure can be predicted from multimodal genomic inputs using an end-to-end convolutional–Transformer architecture. Incorporating DNA methylation yields improvements in both error-based and correlation-based metrics, highlighting its value as an additional epigenomic modality for modeling chromatin organization.

Importantly, our study isolates the contribution of methylation through a controlled ablation framework, indicating that its predictive signal is not fully redundant with chromatin accessibility or CTCF binding. Models trained without methylation tend to produce smoother predictions, whereas methylation-aware models better preserve contrast and structural variation in contact maps.

While our work focuses on local Hi-C prediction and a limited set of evaluation metrics (MSE and Pearson correlation), it provides evidence that DNA methylation contributes meaningfully to learned representations of genome organization. Future work could extend this framework to larger genomic windows, additional cell types, or additional evaluation metrics to further characterize how methylation interacts with other epigenomic signals in shaping 3D chromatin architecture.

4.1 Limitations and Evaluation Scope

While our results demonstrate that multimodal integration improves reconstruction of local Hi-C contact maps, it is important to clarify the scope and limitations of our evaluation. In particular, due to time and computational resource constraints, we did not assess performance using higher-level structural metrics commonly employed in prior 3D genomics studies, such as insulation score correlation, topologically associating domain (TAD) boundary recovery, or chromatin loop detection.

Previous works, including C.Origami and related sequence-based models, emphasize structure-aware evaluation metrics that explicitly quantify domain insulation, loop formation, and distance-stratified interaction patterns. These metrics require extensive post-processing, normalization, and additional computational pipelines, as well as substantial experimental ground truth data. As a result, we focused our evaluation on direct reconstruction-based metrics, including Pearson correlation and mean squared error, computed on predicted Hi-C contact values.

The goal of this study is solely to determine whether incorporating DNA methylation improves predictive performance within a fixed modeling framework. Accordingly, we evaluate performance using Pearson correlation and mean squared error, which provide a consistent basis for comparing models with and without methylation while holding all other factors constant.

Our results are therefore complementary to prior work such as C.Origami, which emphasizes biologically informed structural metrics. Extending our evaluation to include insulation scores, loop-level analyses, and additional structure-aware metrics remains an important direction for future work.