Epigenomics-Guided Flow Matching for 3D Genome Super-Resolution

Yubiao Zhao[1] **Shicheng Song**[2] **Yusen Ye**[3] **Huan Liu**[2] **Juan Liu**[4,*] **Lihua Zhang**[4,*]
[1]School of Cyber Science and Engineering, Wuhan University, Wuhan, China
[2]State Key Laboratory of Oral & Maxillofacial Reconstruction and Regeneration,
Key Laboratory of Oral Biomedicine Ministry of Education,
Hubei Key Laboratory of Stomatology,
School & Hospital of Stomatology, Wuhan University, Wuhan, China
[3]School of Computer Science and Technology, Xidian University, Xi'an, Shaanxi, China
[4]School of Artificial Intelligence, Wuhan University, Wuhan, China
ysye@xidian.edu.cn, liu.huan@whu.edu.cn, {liujuan,zhanglh}@whu.edu.cn

Abstract

Mapping fine-scale 3D chromatin landscapes is critical for understanding gene regulation, yet the resolution of Hi-C data remains fundamentally constrained by both physical bottlenecks and sequencing costs. Although deep learning has shown promise in super-resolving Hi-C data, existing methods suffer from absolute coordinate loss during localized patching and a lack of structural and epigenomic priors in unconditioned generation. We propose a novel epigenomics-guided flow matching framework EpiFlow for 3D genome super-resolution. EpiFlow introduces two key innovations: (1) an EpiCond module that deterministically maps 1D epigenomic features to 2D spatial constraints via pairwise concatenation with built-in positional grounding, replacing cross-attention with a lightweight $\mathcal{O}(L^2)$ fusion pathway; and (2) a Toeplitz-regularized classifier-free guidance mechanism that enforces distance-decay priors using a learnable symmetric prototype matrix. Extensive evaluation shows that our method achieves state-of-the-art cross-cell-type loop detection, improving recall by $2.8\times$ over raw Hi-C on unseen H1-hESC (Loop F_1 of 0.325 vs. 0.199) and outperforming all baselines on same-cell HFFc6 ($F_1 = 0.520$). Aggregate Peak Analysis (APA) confirms biological validity. The model captures transferable features of 3D genome organization, though zero-shot cross-cell transfer remains limited; a lightweight indirect adaptation largely closes this gap, restoring epigenome-only performance to near-native levels (0.021→0.120 vs. native 0.121) and raising both-modality F_1 from 0.325 to 0.352.

1 Introduction

The three-dimensional (3D) folding of chromatin, such as enhancer-promoter loops, is critical for understanding gene regulation and cell identity determination[1]. For over a decade, high-throughput chromosome conformation capture (Hi-C) has served as the dominant technique for mapping these interaction landscapes[2]. However, the resolution of Hi-C is fundamentally restricted by a physical bottleneck: the genomic distribution density of restriction enzyme cleavage sites. Once sequencing depth reaches saturation, the marginal benefit of generating additional data to resolve fine-scale regulatory interactions (e.g., 5 kb or finer) diminishes significantly[3]. To overcome this theoretical limit, alternative technologies such as Micro-C have been developed to achieve nucleosome-level resolution, yet their prohibitive sequencing costs severely restrict large-scale applications[4, 5].

*Corresponding authors.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

Developing computational methods that leap from standard-depth Hi-C data to ultra-high-resolution (e.g., Micro-C level) landscapes offers immense scientific and practical value. Micro-C sequencing costs 5–10$\times$ more per sample than standard-depth Hi-C, posing a major barrier to large-scale application. A method that reliably recovers Micro-C-quality contacts from Hi-C data would dramatically reduce this cost. Deep learning has advanced Hi-C super-resolution. Early "pure vision" approaches treating contact matrices as 2D images[6, 7] use patching strategies that discard absolute genomic coordinates, losing crucial 1D epigenetic guidance[3]. Recent cross-modal approaches like CAESAR[8] integrate these signals via GCNs, but their deterministic architectures and high computational cost restrict them to small sliding windows, limiting long-range topological structures discovery.

Generative models (e.g., DDPMs[9]) have shown promise in 3D genomics[10], yet multimodal conditional guidance remains underexplored. Naively adapting Classifier-Free Guidance (CFG[11]) by replacing conditional signals with zeros causes a fundamental structural inconsistency: it ignores the well-known distance-decay property of contact frequencies, leading to structural "hallucinations"[12]. To resolve this, we go beyond standard zero-replacement and introduce a learnable, symmetric Toeplitz prototype matrix as a distance-aware CFG baseline.

In this study, we propose EpiFlow, a novel generative framework for 3D genome super-resolution based on continuous-time Flow Matching and a lightweight conditional U-Net architecture[13, 14, 15]. Its `EpiCond` module performs pairwise concatenation of 1D epigenetic signals to deterministically map each genomic locus pair to an absolute-coordinate 2D feature vector, while a gated fusion pathway keeps cross-modal fusion at $\mathcal{O}(L^2)$ complexity, avoiding the prohibitive $\mathcal{O}(L^3)$ cost of cross-attention within the fusion module. To enforce structural consistency, we abandon zero-replacement strategies and introduce a learnable prototype vector with symmetric Toeplitz properties. This prototype serves as the distance-decay baseline in CFG inference, maintaining a diagonal-consistent distance-decay profile across any generation path.

In summary, our core contributions are threefold:

- **An Effective Flow Matching Enhancement Framework:** A continuous-time flow matching method with a lightweight conditional U-Net that constructs optimal transport paths via ordinary differential equations (ODEs), enabling high-fidelity generation from standard Hi-C to Micro-C distributions.
- **Cross-Dimensional Alignment and Distance-Aware CFG:** The `EpiCond` module performs pairwise concatenation of 1D epigenetic signals to deterministically anchor pairwise 2D features to absolute genomic coordinates, replacing $\mathcal{O}(L^3)$ cross-attention within the fusion module with a lightweight $\mathcal{O}(L^2)$ fusion pathway. A symmetric Toeplitz prototype serves as the CFG distance-decay baseline, directly encoding chromatin's contact decay profile.
- **Biologically Validated Super-Resolution:** Our method recovers fine-scale topological features unobservable in standard Hi-C, achieving state-of-the-art loop F_1 scores (0.520 on HFFc6; 0.325 on zero-shot cross-cell H1-hESC, rising to 0.352 after indirect adaptation) with APA-confirmed structural fidelity, demonstrating authentic biophysical interactions rather than mathematical artifacts.

2 Related Work

Deep learning for the super-resolution reconstruction of chromosome conformation capture data (e.g., Hi-C and Micro-C) has recently become a central focus in 3D genomics. However, despite progress in data fidelity, existing methods suffer from notable limitations in modeling paradigms, global cross-modal feature mapping, and the handling of distance-dependent structural constraints during generation.

Epigenomics-free Enhancement. Early work framed contact matrix enhancement as a 2D image super-resolution task. For example, HiCPlus [6] used deep convolutional neural networks (CNNs) to infer high-resolution features, but its reliance on Mean Squared Error (MSE) often resulted in overly smoothed predictions. DeepHiC [7] adopted Generative Adversarial Networks (GANs) and perceptual losses to recover high-frequency features, while iEnhance [16] and HiCMamba [17] incorporated multi-scale attention and State Space Models (SSMs) to strengthen spatial dependencies. However, when scaling to genome-wide data, these "pure vision" paradigms heavily rely on localized

patching strategies that sever patches from their absolute genomic coordinates, leaving the models blind to critical biological "signposts," such as epigenetic markers that drive chromatin folding.

Epigenomic Priors and Limitations of Deterministic Regression. Some works have shifted toward integrating 1D epigenomic signals. For instance, CAESAR [8] combined 1D epigenomic signals with low-resolution Micro-C contact maps via a Graph Convolutional Network (GCN) for fine-grained interaction inference. However, such cross-modal methods predominantly adopt deterministic regression architectures. Limited by high computational cost, they are confined to tiny sliding windows (e.g., a 250 kb), restricting the model's capacity to discover novel long-range topological structures at the genome-wide scale. More critically, deterministic models perform point-to-point interpolation, failing to capture the complex, multimodal stochastic distributions underlying chromatin folding.

Generative Models and the Absence of Structural Priors. Generative models have recently emerged in 3D genomics. HiC2MicroC [10] first used Denoising Diffusion Probabilistic Models (DDPMs) to recover Micro-C-specific topological features. However, it treats super-resolution task as a unimodal image-to-image translation, discarding the rich 1D epigenomic context that guides chromatin folding in vivo. As early generative exploations, existing methods lack deep investigation of conditional guidance and structural constraints in multimodal settings. When implementing condition decoupling in generative models (e.g., Classifier-Free Guidance, CFG), naively applying standard computer vision strategies—namely, replacing conditional signals with all-zero tensors—introduces a structurally inconsistent bias. Chromatin contact frequencies exhibit a distance-decay profile: contact probability decreases with genomic distance. Unconditioned inference without this constraint induces structural "hallucinations" that violate this well-established property.

3 Methods

Our goal is to learn a conditional generative mapping from standard-depth Hi-C to ultra-high-resolution Micro-C landscapes. This section presents our epigenomic feature alignment, the continuous-time flow matching framework, and architecture designed to enforce biological structure and distance-dependent consistency. An overview of our proposed framework is illustrated in Figure 1.

3.1 Problem formulation and epigenomic feature alignment

The core objective of our framework is to learn a conditional generative mapping $p_\theta(Y|X, E)$ that transforms a low-resolution Hi-C contact map $X \in \mathbb{R}^{H\times W}$ into a high-fidelity Micro-C landscape $Y \in \mathbb{R}^{H\times W}$, guided by 1D epigenomic signposts $F \in \mathbb{R}^{C\times L}$. Given the extreme heavy-tailed distribution of chromatin contact frequencies, directly feeding raw counts into a generative model leads to numerical instability. Therefore, we first apply a non-linear log-scaling transformation to the raw frequencies c:

$$c' = 2 \cdot \log_{10}\left(\frac{9}{c_{\max}} \cdot \max(0, \min(c, c_{\max})) + 1\right) - 1 \tag{1}$$

This mapping compresses the heavy-tailed distribution into a stable $[-1, 1]$ interval while retaining sufficient dynamic range to distinguish critical fine-scale structures. The clipping threshold $c_{\max} = 0.1$ retains over 99% of valid contacts while suppressing extreme outliers that would otherwise destabilize Flow Matching training.

Biophysically-informed feature alignment. To align heterogeneous 1D epigenomic features and 2D contact map within a unified tensor space, we implement a multi-stage alignment strategy. Initially, raw epigenomic signals at 1bp resolution are downsampled to a 100bp bin size via differential pooling. Specifically, sharp positional signals (e.g., CTCF binding and ATAC-seq) undergo *Max Pooling* to preserve regulatory core intensity, while broader histone modification signals are processed via *Mean Pooling* to reflect regional abundance. Subsequently, these 100bp-resolution signals are fed into a 1D convolutional compression module within the neural network. By employing a cumulative stride of 50 through multiple layers, the signal resolution is precisely mapped to 5kb, achieving absolute coordinate alignment with the contact matrix bins.

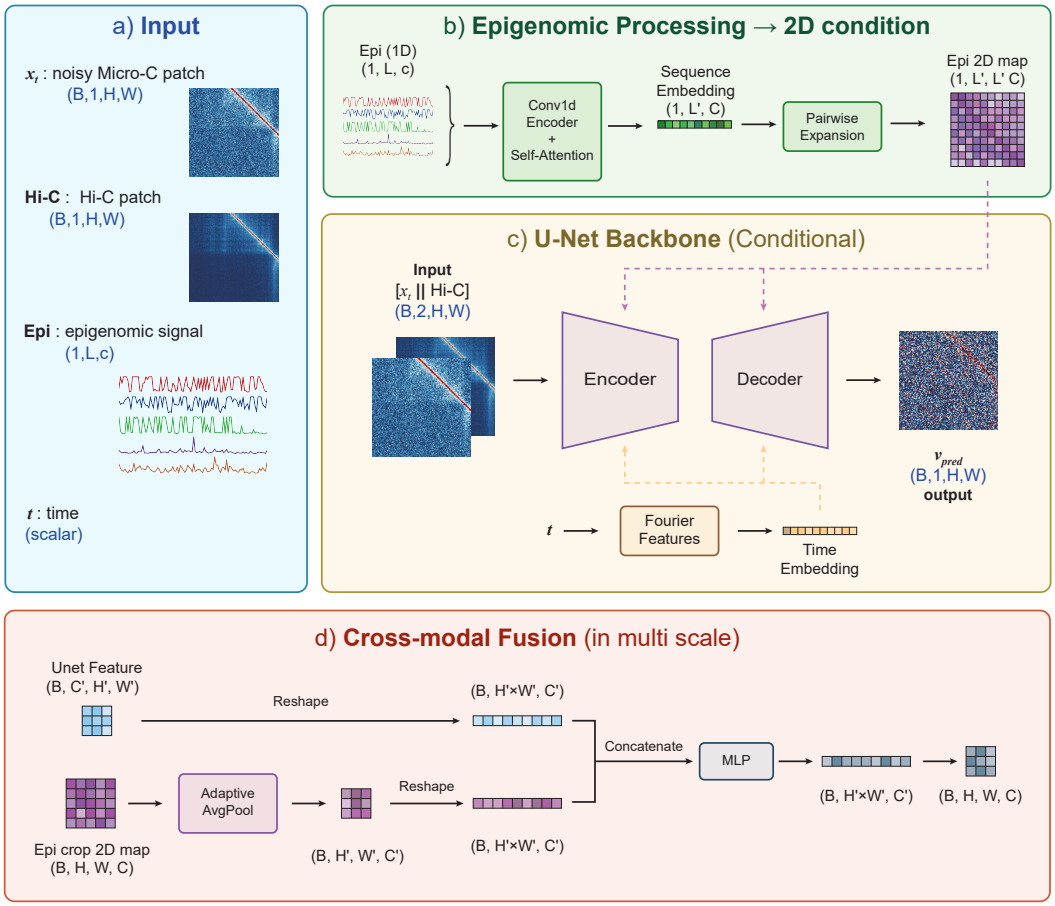

Figure 1: **Overall architecture of the proposed continuous-time multimodal Flow Matching framework. a)** *Inputs*: The model takes the noisy Micro-C patch x_t, low-resolution Hi-C backbone, 1D epigenomic signals, and time step t as multimodal inputs. **b)** *Epigenomic Processing*: 1D signals are encoded via a Conv1D-Attention block and expanded into a 2D structured spatial prior. **c)** *U-Net Backbone*: A conditional U-Net predicts the vector field v_{pred} by integrating the Hi-C backbone and time embeddings. **d)** *Cross-modal Fusion*: A detailed view of the multi-scale fusion mechanism, which employs adaptive pooling and a non-linear MLP projection (instead of heavy cross-attention) to efficiently interweave epigenomic constraints with spatial features.

3.2 Continuous-time flow matching

With the modalities aligned and normalized, we frame the generative task within a continuous-time Flow Matching (FM) paradigm [13, 14]. The FM model constructs a probability density path $p_t(x)$ from a tractable Gaussian prior $x_0 \sim \mathcal{N}(0, I)$ to the target Micro-C distribution x_1 via an Ordinary Differential Equation (ODE):

$$\frac{d}{dt}x_t = u_t(x_t) \tag{2}$$

where u_t is the time-dependent vector field. Following the Optimal Transport (OT) displacement theory, the conditional probability path is defined simply as $x_t = tx_1 + (1-t)x_0$. Consequently, the target vector field that generates this exact path simplifies to $u_t(x_t|x_1) = x_1 - x_0$. We parameterize a multimodal neural network $v_\theta(x_t, t, X, E)$ to approximate this target vector field at any given time step $t \in [0, 1]$.

3.3 Network architecture, training objective, and biological fidelity

Architecture. We employ a parameter-efficient shallow U-Net backbone with four progressively scaled encoding and decoding stages. Self-attention is applied at the two deepest stages, while gated pairwise fusion from the epigenomic prior is applied at all four stages. Continuous time step embeddings are injected via Adaptive Group Normalization (AdaGN), and all convolutional layers use zero-initialized final projection layers for stable training. The companion `EpiCond` encoder contains fewer than 0.1% of the total parameters. Full architectural specifications and parameter counts are provided in Appendix A.3.

Training Objective. We supervise the network v_θ using a Huber loss to ensure robustness against the heavy-tailed outliers characteristic of chromatin contact data:

$$\mathcal{L}_{\text{FM}}(\theta) = \mathbb{E}_{t\sim\mathcal{U}(0,1),x_0,x_1}\left[\mathcal{H}_\delta\Big(v_\theta(x_t,t,X,E) - (x_1 - x_0)\Big)\right] \tag{3}$$

where $x_1 - x_0$ is the optimal transport target. We optimize with AdamW using a linear warmup followed by cosine annealing; complete optimizer settings, learning rate schedule, and training hyperparameters are detailed in Appendix A.3.

Biological Fidelity via Masking. To handle unmapped regions and sequencing artifacts (e.g., repetitive sequences, telomeric "blind spots"), we define a joint validity mask $M_{mask} \in \{0,1\}^{H\times W}$ from regions invalid in either Hi-C or Micro-C. Invalid bins are set to -1 in both modalities. Although our normalization also maps valid zero-contact bins to -1, the network distinguishes them: genuine sparsity appears as localized discrete points, while sequencing blind spots manifest as continuous horizontal and vertical striations. During inference, regions invalid in the input Hi-C are masked to prevent structurally unsupported predictions.

3.4 EpiCond module and efficient multi-modal fusion mechanism

The network v_θ simultaneously processes the noisy state x_t, the low-resolution Hi-C backbone X, and 1D epigenomic markers E. To achieve absolute coordinate alignment and efficient cross-modal fusion within localized patches, we design the following mechanisms.

Precise 1D-to-2D Coordinate Anchoring. The core of EpiCond lies in translating 1D regulatory signals into 2D spatial constraints with built-in positional grounding. First, 1D epigenomic signals E are compressed into strictly one-dimensional hidden states $H_{\text{enc}} \in \mathbb{R}^{C'\times L}$ via a convolutional-attention block with sinusoidal positional encodings, such that $H_{\text{enc},i}$ carries the regulatory signature anchored at absolute genomic coordinate i. We then perform pairwise concatenation of the features at positions i and j:

$$M_{i,j} = \text{Concat}(H_{\text{enc},i}, H_{\text{enc},j}) \tag{4}$$

This construction explicitly embeds the epigenetic context of *both* interacting loci into every entry (i,j) of the 2D feature map. Unlike cross-attention, where the model must learn to associate distant positions through learned queries and keys, the pairwise concatenation guarantees that the feature at (i,j) is deterministically tied to the regulatory signatures at coordinates i and j—achieving precise coordinate anchoring without learned routing. A linear projection followed by a dynamic *Spatial Gate* $G \in (0,1)^{L\times L}$ then sculpts which pairwise interactions are biologically plausible:

$$M' = (G \odot \text{Linear}(M)) + M \tag{5}$$

After forced symmetrization, $M_{\text{sym}} = \frac{1}{2}(M' + M'^{\top})$, this yields a structured 2D spatial prior grounded in absolute coordinates.

Hierarchical Fusion via Gated Projections. Standard cross-attention between the U-Net features (600×600 tokens per window) and epigenomic representations incurs $\mathcal{O}(L^3)$ complexity per head, which is prohibitive at genome scale. We instead propose a multi-stage fusion pathway that is a lightweight gated MLP-based pairwise fusion (not cross-attention) and achieves $\mathcal{O}(L^2)$ complexity; this comparison is scoped to the fusion module. Let F_{spatial} denote the backbone features and F_{epi} the 2D epigenomic prior. We first perform spatial alignment via adaptive pooling to a fixed resolution (64×64), followed by channel-wise concatenation and an MLP projection:

$$F_{align} = \text{MLP}([F_{spatial} \,||\, \text{AdaptiveAvgPool}(F_{epi})]) \tag{6}$$

The aligned features are then refined through a Gated Linear Unit (GEGLU) block to enhance cross-modal interaction:

$$F_{\text{fused}} = \text{Linear}\left(\text{GEGLU}(F_{\text{align}})\right) \tag{7}$$

Finally, we utilize a zero-initialized convolution layer to inject the fused multimodal knowledge back into the main residual branch. This lightweight mechanism enables the model to reconstruct high-fidelity chromatin loops guided by regulatory signals.

3.5 Toeplitz-regularized classifier-free guidance and flexible inference

To enable robust inference under varying conditional strengths, we introduce a Toeplitz-regularized Classifier-Free Guidance (CFG) mechanism. Traditional CFG substitutes dropped conditions with zero tensors. In our normalized contact space, a zero tensor represents a uniform minimum-contact prior ($c' = 0$ after transformation corresponds to $c \to 0$)—a structurally incoherent state where no contacts exist at any genomic distance, directly contradicting the well-established distance-decay property of chromatin. The Toeplitz prototype resolves this conflict by providing a structurally coherent dropout target that respects the characteristic distance-decay profile.

Toeplitz Distance-Decay Prototype. We abandon zero-replacement and introduce a learnable, symmetric Toeplitz prototype matrix $T \in \mathbb{R}^{H \times W}$. We parameterize T with a learnable 1D distance-bias vector w indexed by relative distances $|i - j|$, such that $T_{i,j} = w_{|i-j|}$. During training, Hi-C and epigenomic conditions are randomly dropped (see Appendix A.3 for rates). When conditions are removed, the network adaptively falls back to T, ensuring the generative path respects the characteristic distance-decay profile. Figure 1 and Appendix E.3 visualize and analyze the learned prototype parameters.

Discrete Guidance and High-Order Sampling. In the inference phase, we control the injection of multimodal constraints using a guidance scale $\omega \in \{1.0, 1.5\}$. Let $v_\theta(x_t|\emptyset)$ denote the unconditioned prediction using the Toeplitz prototype. The calibrated vector field is defined as:

$$\tilde{v}_t = v_\theta(x_t|\emptyset) + \omega \cdot \big(v_\theta(x_t|C) - v_\theta(x_t|\emptyset)\big) \tag{8}$$

where $\omega = 1.0$ represents standard conditional inference and $\omega = 1.5$ enhances the influence of multimodal signposts C.

To solve the continuous-time ODE defined by $\tilde{v}_t$, we employ Heun's Second-Order Method with 20 uniformly spaced time steps ($\Delta t = 1/20$). At each step, the solver evaluates $v_1 = \tilde{v}_t(x_t)$, computes an intermediate state $\tilde{x} = x_t + v_1 \Delta t$, evaluates the correction $v_2 = \tilde{v}_t(\tilde{x})$, and updates the trajectory with the averaged gradient: $x_{t+\Delta t} = x_t + \frac{1}{2}(v_1 + v_2)\Delta t$[18].

4 Experiments

In this section, we evaluate our EpiFlow on the 3D genome super-resolution task. We assess three key aspects: (1) cross-cell-type generalization capability across different cell lines; (2) the impact of Classifier-Free Guidance (CFG) on multi-scale structural restoration; and (3) the individual contributions of input modalities and the spatial gating mechanism.

4.1 Experimental setup

Datasets and Data Partitioning. We use autosomal sequencing data of the HFFc6 and H1-hESC cell lines from 4DN[19] and ENCODE[20]. The resolution for both input Hi-C and target Micro-C matrices is set to 5 kb with a 3 Mb genomic window. For the baseline comparisons, DeepHiC was retrained from scratch on the HFFc6 dataset using its publicly available code and default hyperparameters. For HiC2MicroC, since its original publication already utilized HFFc6 as the training set, we directly evaluated it using the pre-trained weights and inference scripts provided in their official repository. This ensures all methods are fairly evaluated on the same target domain. To evaluate the model's cross-cell-type generalizability, we adopt a strict transfer learning setting:

- **Training Set:** Autosomes of the HFFc6 cell line, with chromosomes 8, 9, 11, and 22 strictly excluded to prevent structural information leakage.
- **Intra-Cell-Type Test Set:** The held-out chromosomes 8, 9, 11, and 22 of the HFFc6 cell line.

- **Cross-Cell-Type Test Set:** Chromosomes 8, 9, 11, and 22 of the unseen H1-hESC cell line. In this setting, the model relies entirely on the weights learned from HFFc6 without any fine-tuning.

The 1D epigenomic signposts consist of five key biomarkers: CTCF, ATAC, H3K4me3, H3K27ac, and H3K4me1. During the training phase, we employ a dynamic random cropping strategy to increase sample diversity and mitigate the structural rigidity induced by fixed-step windows.

Evaluation Protocol. To ensure standardized biological evaluation, all predicted tensors are inverse-transformed to raw contact frequencies and encapsulated into the standard `.cool` format [21]. Numerical consistency is assessed using the Pearson Correlation Coefficient (PCC), Spearman Rank Correlation Coefficient (SRCC), Structural Similarity Index (SSIM), and Stratum-adjusted Correlation Coefficient (SCC) [22]. Discrete structural discovery is evaluated via F_1 scores for Chromatin Loops (1-bin tolerance, detected by Mustache [23]) and Topologically Associating Domains (TADs, 50 kb window, detected by Cooltools[24]).

4.2 Main results: image fidelity and cross-cell-type structure discovery

This section presents a comprehensive quantitative comparison between our proposed framework and current baseline models. To rigorously validate cross-cell-type generalization, all test results for both EpiFlow and the baselines on the H1-hESC cell line were generated using weights trained exclusively on the HFFc6 cell line (i.e., pure cross-cell-type inference). Table 1 summarizes the continuous numerical and distance-decay correlation metrics, while Table 2 highlights the discovery rates of discrete authentic biological structures (e.g., Loops and TADs).

Table 1: Quantitative comparison of continuous numerical fidelity and structural consistency. The best results among the generative models are highlighted in **bold**, and the second best are underlined. *Evaluated in a strict cross-cell-type setting using HFFc6-trained weights. For EpiFlow, CFG denotes the Classifier-Free Guidance scale, where CFG 1.0 represents standard conditional generation and CFG 1.5 applies a moderate guidance scale to enhance conditioning.

Cell Type	Method	PCC ↑	SRCC ↑	SSIM ↑	SCC ↑
HFFc6	Raw Hi-C	0.9150	0.8047	0.9634	0.7112
	DeepHiC	0.9967	**0.8612**	0.9888	**0.8926**
	HiC2MicroC	0.9872	0.8477	0.9832	0.8515
	EpiFlow (CFG 1.0)	**0.9974**	0.8461	**0.9905**	0.7867
	EpiFlow (CFG 1.5)	**0.9974**	0.8466	0.9902	0.8058
H1-hESC*	Raw Hi-C	0.8573	0.7747	0.8949	0.4280
	DeepHiC	0.9719	**0.8615**	0.9181	0.6114
	HiC2MicroC	**0.9889**	0.8150	0.9068	**0.6142**
	EpiFlow (CFG 1.0)	0.9776	0.8422	**0.9479**	0.6065
	EpiFlow (CFG 1.5)	0.9625	0.8443	0.9369	0.5465

Image-Level Fidelity and the "Oversmoothing" Illusion of SCC. In metrics reflecting pixel-level spatial mapping and texture restoration, our model demonstrates a consistent advantage. Particularly in the Structural Similarity Index (SSIM), our model (CFG 1.0) achieves an optimal 0.9905 in the intra-cell-type HFFc6 test and maintains a state-of-the-art 0.9479 in the cross-cell-type H1-hESC test. In contrast, while baseline models such as traditional GAN architectures (DeepHiC) and existing diffusion models (HiC2MicroC) manage to maintain relatively high SSIM scores, they experience more noticeable degradation during cross-cell-type generalization (dropping to 0.9181 and 0.9068 on H1-hESC, respectively), falling behind our multimodal framework.

However, as observed in Table 1, baseline models achieve higher SCC (DeepHiC: 0.893 on HFFc6) while SRCC values remain comparable (within 0.05). SCC penalizes per-stratum variance, rewarding oversmoothed matrices that suppress biological variability; rank-based SRCC is insensitive to this compression. This divergence provides a practical diagnostic: if a method achieves high SCC without a corresponding SRCC gain, it is likely oversmoothing. P(s) and per-stratum SCC analysis

(Appendix E.1 and E.2) confirm that baselines systematically deviate from the Micro-C decay profile and inflate per-stratum correlations, while our method closely tracks the true distance-decay profile ($r = 0.995$).

Table 2: Quantitative comparison of discrete biological structure discovery rates. The best results among the generative models are highlighted in **bold**, and the second best are underlined. Our framework maintains robust structural discovery capabilities even in the rigorous cross-cell-type setting. For EpiFlow, CFG denotes the Classifier-Free Guidance scale, where CFG 1.0 represents standard conditional generation and CFG 1.5 applies a moderate guidance scale to enhance conditioning.

Cell Type	Method	TAD F_1 (50kb) ↑	Loop Precision ↑	Loop Recall ↑	Loop F_1 (1-bin) ↑
HFFc6	Raw Hi-C	0.5000	0.6199	0.3455	0.4437
	DeepHiC	0.5538	0.2889	0.5028	0.3669
	HiC2MicroC	0.5690	0.3488	0.6332	0.4498
	EpiFlow (CFG 1.0)	0.5779	**0.4333**	0.6507	**0.5202**
	EpiFlow (CFG 1.5)	**0.6019**	0.3750	**0.6750**	0.4821
H1-hESC*	Raw Hi-C	0.3405	0.5262	0.1225	0.1987
	DeepHiC	**0.4242**	0.1858	0.2247	0.2034
	HiC2MicroC	0.2441	0.2161	0.3546	0.2685
	EpiFlow (CFG 1.0)	0.3446	**0.3115**	0.3404	**0.3253**
	EpiFlow (CFG 1.5)	0.3571	0.2533	**0.4029**	0.3110

Robust Performance in Cross-Cell-Type Structure Discovery. Table 2 reveals an inverse relationship between SCC and loop detection: the highest-SCC methods (DeepHiC, HiC2MicroC) achieve the lowest Loop F_1, consistent with oversmoothing suppressing fine-scale features. Our model achieves a Loop F_1 of 0.5202 (HFFc6), outperforming HiC2MicroC (0.4498), and maintains robust cross-cell performance on H1-hESC ($F_1 = 0.3253$ vs. 0.2685).

Raw Hi-C is highly conservative (precision: 0.620/0.526, recall: 0.346/0.123 on HFFc6/H1-hESC). Our approach increases recall substantially ($1.9\times$–$2.8\times$ relative to Hi-C). Our method also achieves the highest precision among all generative models (0.433/0.312 on HFFc6/H1 vs. DeepHiC's 0.289/0.186 and HiC2MicroC's 0.349/0.216), leading on both precision and recall simultaneously. APA of predicted false positives (Appendix D.2) confirms that FP predictions are not random: they exhibit enrichment above 1.0 ($1.30\times$–$1.39\times$), representing genuine 3D contacts below the loop-caller detection threshold. At higher tolerances ($\tau \to 3$), the precision gap between our method and Hi-C narrows substantially. This validates our design: unlike methods that globally smooth to inflate SCC, our precise coordinate anchoring recovers fine-scale structures at specific genomic loci rather than memorizing cell-specific texture.

Visual inspection of generated contact maps (Appendix C) reveals weaknesses invisible to aggregate metrics: HiC2MicroC produces unnaturally stark contrast, while DeepHiC exhibits irregular diagonal streaks and moiré-like textures absent from real Micro-C. Our method faithfully reconstructs TAD boundaries and punctate loops with natural texture.

Multi-scale Structural Trade-off via CFG. Comparing CFG scales in Table 2 reveals a tunable trade-off: CFG 1.0 favors precision and overall F_1, while CFG 1.5 boosts recall (0.3404→0.4029 on H1-hESC) and TAD F_1 (0.3446→0.3571) at the cost of precision—benefiting domain-boundary detection and long-range loop discovery. DeepHiC achieves the highest TAD F_1 on H1-hESC (0.4242) via global smoothing, beneficial for broad domains but detrimental to fine-scale loops.

4.3 Cross-cell-type adaptation via indirect transfer

Pure cross-cell-type inference leaves a gap on H1-hESC ($F_1 = 0.3253$ with both modalities). We therefore evaluate an indirect adaptation strategy that avoids full retraining: pseudo Micro-C labels for H1-hESC are first generated by the HFFc6-trained model under Hi-C-only conditioning (the epigenomic branch is dropped and replaced by the learned Toeplitz prototype); these labels then supervise adaptation of EpiCond, the fusion layers, and the related prototypes with the U-Net backbone frozen, using H1-hESC Hi-C and epigenomic inputs. As shown in Table 3, indirect

adaptation improves over direct transfer under Hi-C + epigenomic conditioning (0.3517 vs. 0.3253) and substantially recovers the epigenomic-only setting (0.0207 $\rightarrow$ 0.1196), which is comparable to a model trained natively on H1-hESC (0.1209). These results indicate that epigenomic conditioning becomes informative after appropriate adaptation even though no real Micro-C labels from the target cell type are used—the adaptation is supervised entirely by pseudo-labels generated by the source-trained model—and the adapted model still achieves solid performance. Direct, zero-shot transferability of the epigenomic branch remains limited, most likely due to epigenomic batch effects across cell types; the adaptation above largely mitigates this limitation.

Table 3: Direct transfer vs. module fine-tuning on H1-hESC (loop F_1, CFG scale 1.0). Indirect adaptation updates EpiCond, the fusion layers, and the related prototypes with the U-Net backbone frozen. Best results in **bold**.

Method	Both	Hi-C only	Epi only
Direct transfer (HFFc6→H1-hESC)	0.3253	**0.3197**	0.0207
Indirect transfer (HFFc6→H1-hESC)	**0.3517**	0.3092	**0.1196**
Native train on H1-hESC (reference)	0.4037	0.3281	0.1209

4.4 Ablation: multi-modal synergy and spatial gating

To investigate the contribution of individual modalities and the effectiveness of the `EpiCond` module, we conduct an ablation study on the H1-hESC cell line. Unlike the main cross-cell-type evaluation, the ablation uses an intra-cell-type setting (trained and tested on H1-hESC) to isolate the effects of individual architectural components without the confounding factor of cell-type transfer. We independently evaluate the 1D epigenomic markers (Epi) and the 2D contact map (Hi-C), alongside a comparison of feature fusion methods (gated vs. ungated). The quantitative results are summarized in Table 4.

Table 4: Ablation study on input modalities and the spatial gating mechanism (evaluated on H1-hESC intra-cell, 1-bin tolerance). Single-modality models use the gated architecture for fair comparison with the full model. The best results are highlighted in **bold**, and the second best are underlined.

Model Variant	Epi	Hi-C	Gated	Precision ↑	Recall ↑	Loop F_1 ↑
Raw Hi-C	–	–	–	0.5262	0.1225	0.1987
Epi Only (Gated)	✓	×	✓	0.0947	0.1674	0.1209
Hi-C Only (Gated)	×	✓	✓	0.2554	0.4586	0.3281
Both Modalities (Ungated)	✓	✓	×	<u>0.2931</u>	<u>0.5527</u>	<u>0.3831</u>
Both Modalities (Gated)	✓	✓	✓	**0.3057**	**0.5942**	**0.4037**

Modality Contributions. As shown in Table 4, the Hi-C-only model achieves a Loop F_1 of 0.3281, confirming that 2D contact maps provide the fundamental structural scaffold. Conditioning solely on 1D Epi signals yields $F_1 = 0.1209$, verifying the model's ability to recover partial topological structures *de novo* from regulatory marks alone. The integration of both modalities reaches $F_1 = 0.4037$, demonstrating that epigenomic signals effectively guide fine-scale structural refinement upon the contact framework provided by Hi-C.

Effectiveness of Spatial Gating. On the full dual-modality model, the spatial gate improves F_1 from 0.3831 to 0.4037, driven primarily by a recall increase (0.5527→0.5942) with a smaller precision gain (0.2931→0.3057). This indicates that the gate acts as a dynamic filter: it identifies and retains authentic chromatin loops that are obscured in low-resolution data while suppressing cross-modal noise, resulting in sharper topological boundaries without sacrificing specificity. The gate's minimal parameter cost (257 parameters, $< 0.001\%$ of the total) further underscores the efficiency of the design.

5 Discussion

We presented a multimodal Flow Matching framework for 3D genome super-resolution combining cross-dimensional epigenomic alignment (EpiCond) with Toeplitz-regularized CFG. On cross-cell H1-hESC, our method achieves $2.8\times$ the recall of raw Hi-C (F_1=0.325 vs. 0.199); on HFFc6 it reaches F_1=0.520 with $1.9\times$ recall improvement, with APA confirming genuine focal enrichment. Visual inspection further confirms that our reconstructions faithfully preserve natural chromatin texture, avoiding the oversmoothing and striation artifacts observed in existing baselines. CFG scale tunes the precision–recall and loop–TAD trade-offs. A lightweight indirect adaptation largely resolves the cross-cell gap: both-modality F_1 rises from 0.325 to 0.352, and epigenome-only performance recovers to a near-native level (0.120 vs. 0.121; Section 4.3).

A methodological contribution of this work is establishing the SCC-SRCC divergence as a practical diagnostic for generative model evaluation in 3D genomics. Standard SCC penalizes per-stratum variance, rewarding oversmoothed matrices that suppress biological variability; rank-based SRCC is insensitive to this compression. Methods with highest SCC without corresponding SRCC gain (DeepHiC: 0.893 on HFFc6; HiC2MicroC) consistently achieve the lowest loop F_1, while our framework achieves state-of-the-art structural recovery without sacrificing continuous fidelity. Per-stratum decomposition (Appendix E.2) dissects this phenomenon, revealing HiC2MicroC's catastrophic long-range collapse ($\rho_d \approx 0.05$ at >1.5 Mb), which our Toeplitz prototype remedies by encoding the characteristic distance-decay profile directly into the CFG mechanism. From an architectural perspective, EpiCond replaces $\mathcal{O}(L^3)$ cross-attention with $\mathcal{O}(L^2)$ gated fusion within the fusion module, embodying the design principle of embedding structural priors into the architecture rather than relying solely on data-driven learning.

Limitations. (1) Pure epigenomic prediction collapses from F_1=0.121 (same-cell) to F_1=0.023 under direct transfer; the full dual-modality model drops more moderately from F_1=0.404 to F_1=0.325, indicating that the Hi-C backbone provides partial robustness against domain shift. Retraining on the target cell type recovers performance in both cases (Table 4), and an indirect adaptation strategy that fine-tunes only the conditioning modules narrows the gap substantially (Section 4.3; both-modality 0.325 to 0.352; epigenome-only 0.021 to 0.120), confirming that the architecture itself is capable and the gap stems from cell-type-specific feature distributions rather than a fundamental capacity limitation. The direct-transfer gap stems from cell-type-specific epigenomic signatures that do not transfer across lineages; the indirect adaptation above mitigates this (epigenome-only transfer becomes comparable to native training), and additional marks (e.g., cohesin subunits SMC3 or RAD21 ChIP-seq) may further reduce the residual discrepancy. (2) Validation is currently limited to two human cell lines (HFFc6 and H1-hESC); broader evaluation across cell types and species is needed.

Future directions include extension to disease-relevant structural variants, integration with single-cell multi-omics, and spatially varying CFG scales. By establishing that a Toeplitz prototype can encode the distance-decay constraint and that $\mathcal{O}(L^2)$ gated fusion provides a tractable alternative to $\mathcal{O}(L^3)$ cross-attention at genome scale, this work demonstrates that structural priors and efficient multimodal fusion are not competing goals. The framework provides a practical tool for uncovering cis-regulatory mechanisms underlying complex diseases and developmental processes.

Acknowledgments and Disclosure of Funding

This work was supported by the National Key Research and Development Program of China (Grant No. 2023YFF0725400 to L.Z.). The authors would also like to thank the anonymous reviewers for their constructive feedback and suggestions.

References

[1] Job Dekker and Leonid Mirny. The 3D Genome as Moderator of Chromosomal Communication. *Cell*, 164(6):1110–1121, March 2016.

[2] Erez Lieberman-Aiden, Nynke L. van Berkum, Louise Williams, Maxim Imakaev, Tobias Ragoczy, Agnes Telling, Ido Amit, Bryan R. Lajoie, Peter J. Sabo, Michael O. Dorschner, Richard Sandstrom, Bradley Bernstein, M. A. Bender, Mark Groudine, Andreas Gnirke, John Stamatoyannopoulos, Leonid A. Mirny, Eric S. Lander, and Job Dekker. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 326(5950):289–293, 2009.

[3] Suhas S.P. Rao, Miriam H. Huntley, Neva C. Durand, Elena K. Stamenova, Ivan D. Bochkov, James T. Robinson, Adrian L. Sanborn, Ido Machol, Arina D. Omer, Eric S. Lander, and Erez Lieberman Aiden. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell*, 159(7):1665–1680, December 2014.

[4] Tsung-Han S. Hsieh, Assaf Weiner, Bryan Lajoie, Job Dekker, Nir Friedman, and Oliver J. Rando. Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C. *Cell*, 162(1):108–119, July 2015.

[5] Nils Krietenstein, Sameer Abraham, Sergey V. Venev, Nezar Abdennur, Johan Gibcus, Tsung-Han S. Hsieh, Krishna Mohan Parsi, Liyan Yang, René Maehr, Leonid A. Mirny, Job Dekker, and Oliver J. Rando. Ultrastructural Details of Mammalian Chromosome Architecture. *Molecular Cell*, 78(3):554–565.e7, May 2020.

[6] Yan Zhang, Lin An, Jie Xu, Bo Zhang, W. Jim Zheng, Ming Hu, Jijun Tang, and Feng Yue. Enhancing Hi-C data resolution with deep convolutional neural network HiCPlus. *Nature Communications*, 9(1):750, February 2018.

[7] Hao Hong, Shuai Jiang, Hao Li, Guifang Du, Yu Sun, Huan Tao, Cheng Quan, Chenghui Zhao, Ruijiang Li, Wanying Li, Xiaoyao Yin, Yangchen Huang, Cheng Li, Hebing Chen, and Xiaochen Bo. DeepHiC: A generative adversarial network for enhancing Hi-C data resolution. *PLOS Computational Biology*, 16(2):e1007287, February 2020.

[8] Fan Feng, Yuan Yao, Xue Qing David Wang, Xiaotian Zhang, and Jie Liu. Connecting high-resolution 3D chromatin organization with epigenomics. *Nature Communications*, 13(1):2054, April 2022.

[9] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS '20, Red Hook, NY, USA, 2020. Curran Associates Inc.

[10] Tong Liu, Hao Zhu, and Zheng Wang. Learning Micro-C from Hi-C with diffusion models. *PLOS Computational Biology*, 20(5):e1012136, May 2024.

[11] Jonathan Ho and Tim Salimans. Classifier-Free Diffusion Guidance, 2022.

[12] Geoffrey Fudenberg, Maxim Imakaev, Carolyn Lu, Anton Goloborodko, Nezar Abdennur, and Leonid A. Mirny. Formation of Chromosomal Domains by Loop Extrusion. *Cell Reports*, 15(9):2038–2049, May 2016.

[13] Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *The Eleventh International Conference on Learning Representations*, 2023.

[14] Alexander Tong, Kilian FATRAS, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. *Transactions on Machine Learning Research*, 2024.

[15] Katherine Crowson, Stefan Andreas Baumann, Alex Birch, Tanishq Mathew Abraham, Daniel Z Kaplan, and Enrico Shippole. Scalable high-resolution pixel-space image synthesis with hourglass diffusion transformers. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp, editors, *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 9550–9575. PMLR, 21–27 Jul 2024.

[16] Kai Li, Ping Zhang, Zilin Wang, Wei Shen, Weicheng Sun, Jinsheng Xu, Zi Wen, and Li Li. iEnhance: A multi-scale spatial projection encoding network for enhancing chromatin interaction data resolution. *Briefings in Bioinformatics*, 24(4):bbad245, July 2023.

[17] Minghao Yang, Zhi-An Huang, Zhihang Zheng, Yuqiao Liu, Shichen Zhang, Pengfei Zhang, Hui Xiong, and Shaojun Tang. HiCMamba: Enhancing Hi-C resolution and identifying 3D genome structures with state space modeling. *PLOS Computational Biology*, 22(3):e1014057, March 2026.

[18] Tero Karras, Miika Aittala, Samuli Laine, and Timo Aila. Elucidating the design space of diffusion-based generative models. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, NIPS '22, Red Hook, NY, USA, 2022. Curran Associates Inc.

[19] the 4D Nucleome Network, Job Dekker, Andrew S. Belmont, Mitchell Guttman, Victor O. Leshyk, John T. Lis, Stavros Lomvardas, Leonid A. Mirny, Clodagh C. O'Shea, Peter J. Park, Bing Ren, Joan C. Ritland Politz, Jay Shendure, and Sheng Zhong. The 4D nucleome project. *Nature*, 549(7671):219–226, September 2017.

[20] The ENCODE Project Consortium et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature*, 583(7818):699–710, July 2020.

[21] Nezar Abdennur and Leonid A Mirny. Cooler: Scalable storage for Hi-C data and other genomically labeled arrays. *Bioinformatics*, 36(1):311–316, January 2020.

[22] Tao Yang, Feipeng Zhang, Galip Gürkan Yardımcı, Fan Song, Ross C. Hardison, William Stafford Noble, Feng Yue, and Qunhua Li. HiCRep: Assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient. *Genome Research*, 27(11):1939–1949, November 2017.

[23] Abbas Roayaei Ardakany, Halil Tuvan Gezer, Stefano Lonardi, and Ferhat Ay. Mustache: Multi-scale detection of chromatin loops from Hi-C and Micro-C maps using scale-space representation. *Genome Biology*, 21(1):256, December 2020.

[24] Open2C, Nezar Abdennur, Sameer Abraham, Geoffrey Fudenberg, Ilya M. Flyamer, Aleksandra A. Galitsyna, Anton Goloborodko, Maxim Imakaev, Betul A. Oksuz, Sergey V. Venev, and Yao Xiao. Cooltools: Enabling high-resolution Hi-C analysis in Python. *PLOS Computational Biology*, 20(5):e1012067, May 2024.

A Datasets and reproducibility

A.1 Dataset accession and formats

To ensure the strict reproducibility of our study, we detail the provenance and specific file formats of all multimodal datasets used for the intra- and cross-cell-type evaluations. Data were collected for two primary human cell lines: HFFc6 (human foreskin fibroblasts) and H1-hESC (human embryonic stem cells).

Biological Rationale for 1D Epigenomic Signposts. To effectively guide the generative process of 3D chromatin structures, we selected five 1D epigenomic markers—profiled via Chromatin Immunoprecipitation Sequencing (ChIP-seq) and ATAC-seq—that serve as fundamental biophysical "signposts" for chromatin folding and gene regulation:

- **CTCF (ChIP-seq):** A highly conserved architectural protein that acts as the primary anchor for chromatin loops and defining boundaries of Topologically Associating Domains (TADs).
- **ATAC-seq:** Measures general chromatin accessibility, broadly identifying open regulatory regions that are physically available for interactions.
- **H3K27ac (ChIP-seq):** A robust marker for active enhancers and promoters, strongly correlating with active transcriptional interactions.
- **H3K4me1 (ChIP-seq):** Marks primed or active enhancer elements, delineating the regulatory landscape for distal interactions.
- **H3K4me3 (ChIP-seq):** Strictly enriched at active gene promoters, providing precise anchor points for promoter-captured loops.

By integrating these markers, the model learns to associate 1D regulatory signals with 2D spatial folding patterns, using epigenomic features as positional anchors that inform the generative process.

Data Sourcing and Formatting. Both the standard-depth Hi-C inputs and the ultra-high-resolution Micro-C targets were sourced from the 4D Nucleome (4DN) Data Portal. To maintain uniform structural representation, we strictly utilized the processed multi-resolution `.mcool` format files. For epigenomic priors, we curated ChIP-seq and ATAC-seq tracks representing the "Fold Change over Control" (FC) signal, ensuring the Flow Matching framework learns from biologically consistent, normalized intensity values across experiments. Files sourced from ENCODE were downloaded in the `.bigWig` format, while those from 4DN were acquired in the `.bw` format.

The exact file accession identifiers for all modalities and cell lines are comprehensively summarized in Table 5.

Table 5: **Detailed Accession Codes for the Multi-modal Datasets.** All 1D Epigenomic tracks (ChIP-seq and ATAC-seq) represent "Fold Change over Control" signals. The specific file formats and database accession identifiers for both cell lines are listed below.

Data Modality	Format	HFFc6 ID	H1-hESC ID
3D Genomic Contact Matrices (4DN Portal)			
Hi-C	`.mcool`	`4DNFIB59T7NN`	`4DNFI6HDY7WZ`
Micro-C	`.mcool`	`4DNFI9FVHJZQ`	`4DNFI9GMP2J8`
1D Epigenomic Signposts (4DN & ENCODE Portals)			
H3K4me1 (ChIP-seq)	`.bw`	`4DNFI3WBAYI7`	`4DNFIDPEPJ5B`
H3K4me3 (ChIP-seq)	`.bw`	`4DNFIVYWZE1W`	`4DNFIQP1BBB3`
H3K27ac (ChIP-seq)	`.bw`	`4DNFINRI6WOL`	`4DNFIRN137C9`
ATAC-seq	`.bw`	`4DNFIZ9191QU`	`4DNFICPNO4M5`
CTCF (ChIP-seq)	`.bigWig`	`ENCFF951KQM`	`ENCFF038RVZ`

A.2 Data preprocessing pipeline

The multimodal data undergo a standardized preprocessing pipeline to ensure tensor compatibility and numerical stability before entering the Flow Matching framework.

Contact Matrix Normalization. Raw Hi-C and Micro-C contact matrices are first subjected to iterative correction (ICE) to remove locus-specific and experimental biases. After normalization, contact frequencies are clipped to $c_{\max} = 0.1$ and transformed via the log-scaling equation in the main text (Equation 1), which maps valid contacts to $[-1, 1]$. Bins with insufficient sequencing coverage (e.g., centromeric and telomeric regions) are set to -1 and recorded in a binary validity mask.

Epigenomic Signal Alignment. Raw epigenomic signals at 1 bp resolution are first downsampled to $\sim$100 bp bins using modality-specific pooling. Sharp, punctate signals (CTCF ChIP-seq, ATAC-seq) undergo max pooling to preserve local peak intensity, while diffuse histone modification signals (H3K4me1, H3K4me3, H3K27ac) undergo mean pooling to capture regional enrichment. The 100 bp-resolution tracks are then fed into the `EpiCond` module's 1D convolutional compressor, which applies three Conv1d layers with kernel sizes $[11, 11, 5]$ and strides $[5, 5, 2]$, yielding a cumulative stride of 50. This maps the signals precisely from $\sim$100 bp to 5,000 bp (5 kb) bins, achieving absolute coordinate alignment with the contact matrices.

Window Partitioning and Patch Extraction. The genome is partitioned into overlapping 3 Mb windows (600 bins at 5 kb resolution) with a 1 Mb stride. Each window is stored as a separate `.npz` file containing the Hi-C sub-matrix (600×600), Micro-C sub-matrix (600×600), epigenomic tracks ($5 \times 30,000$, covering 100 bp-resolution signals across the 3 Mb window), and genomic coordinates of the window origin. During training, 128×128 sub-patches (640 kb $\times$ 640 kb) are dynamically cropped from random positions within each window, providing continuous spatial augmentation. 16 such patches are sampled per window per training step. During inference, 128×128 patches are extracted with a stride of 118 (10-bin overlap), and overlapping predictions are averaged in the transformed $[-1, 1]$ space before inverse transformation. The final predicted matrix is symmetrized to enforce $M_{i,j} = M_{j,i}$.

A.3 Training and hyperparameter details

To ensure full reproducibility and to provide transparency regarding our experimental setup, we detail the optimization strategies, data sampling, and hyperparameter selection rationale used during the training of our continuous-time Flow Matching framework.

Data Splitting and Spatial Sampling. Training used the autosomes of HFFc6 excluding chr8, 9, 11, and 22 (held out for testing). Dynamic random cropping of 128×128 patches (640 kb $\times$ 640 kb) from whole-genome matrices served as continuous spatial augmentation. No significant training–validation loss gap was observed across 50 epochs, and the final checkpoint was used for all evaluations without early stopping.

Optimization and Hyperparameters. The network was optimized using `AdamW` with initial learning rate 3×10^{-4}, weight decay 1×10^{-3}, and momentum parameters $\beta_1 = 0.9$, $\beta_2 = 0.95$. We used a composite learning rate schedule: one epoch of linear warmup (scaling from $0.01\times$ to $1.0\times$ the base rate) followed by cosine annealing to $\eta_{\min} = 1 \times 10^{-6}$ over the remaining 49 epochs. Training used a batch size of 4 with 128×128 patches and bfloat16 mixed precision. CFG dropout rates (0.3 for Hi-C, 0.2 for epigenomic) were selected via grid search.

Model Architecture and Parameter Count. The complete framework consists of a 4-stage conditional U-Net backbone with channel dimensions $[128, 256, 512, 1024]$ and layer depths $[2, 2, 4, 4]$, totaling **454,981,592** parameters, plus the `EpiCond` encoder with **358,786** parameters. The combined framework contains **455,340,378** trainable parameters ($\sim$455M); the epigenomic encoder accounts for only 0.08% of the total. The ungated ablation variant (`EpiCondModel`) contains 358,529 parameters (257 fewer, corresponding to the spatial gate linear layer). Self-attention is applied at the two deepest U-Net stages, while gated pairwise fusion from the epigenomic prior is applied at all four stages. All convolutional layers use 3×3 kernels with zero-initialized final projection layers.

A.4 Computational resources

All experiments, including model training, validation, and inference, were conducted on a Linux-based high-performance computing (HPC) server. The detailed hardware specifications and software configurations are as follows:

Hardware Infrastructure. The physical host node is equipped with a dual-socket Intel(R) Xeon(R) Platinum 8269CY @ 2.50GHz CPU (yielding a total of 104 logical threads) and 1.0 TiB of system memory. The server contains a GPU cluster of 9 $\times$ NVIDIA GeForce RTX 3090 GPUs.

Single-GPU Execution Strategy and Computational Cost. Although the host server is equipped with a multi-GPU cluster, we emphasize that our proposed continuous-time Flow Matching framework is highly parameter-efficient. **All individual training runs, validations, and genome-wide inferences in this study were strictly executed on a single NVIDIA RTX 3090 GPU (24 GB VRAM).** Empirically, completing a full training cycle of 50 epochs requires approximately 40 to 50 hours of continuous execution on a single card. The multi-GPU environment was solely utilized to execute multiple independent experimental trials (e.g., hyperparameter tuning and cross-cell-line evaluations) concurrently.

Software Environment. The computational framework was developed using Python 3.11.11. The deep learning architectures and Flow Matching ODE solvers were implemented in PyTorch (version 2.5.0) compiled with CUDA 12.4 for optimal GPU acceleration. For 3D genomic data processing, matrix manipulation, and discrete structural evaluation, we extensively utilized standard bioinformatics libraries, including `cooler` for multi-resolution data extraction, `cooltools` for Topologically Associating Domain (TAD) calling, and `mustache` for chromatin loop detection.

B Evaluation metrics details

To rigorously assess the fidelity of the generated 3D chromatin structures, we align our evaluation metrics with the narrative progression of the main text: first examining the continuous spatial correlations and pixel-level similarity, and subsequently evaluating the discovery rate of discrete biological features. Here, we detail the mathematical formulations and computational implementations for these metrics.

Continuous Numerical Fidelity and Pixel-Level Metrics. To evaluate the pixel-level texture preservation, we calculated the Structural Similarity Index (SSIM), Pearson Correlation Coefficient (PCC), and Spearman Rank Correlation Coefficient (SRCC). Computing global metrics directly on the entire chromosomal matrix can heavily bias the results due to unmappable regions. Therefore, we implemented a rigorous sliding-window evaluation protocol:

1. **Strict Block Validation:** We traversed the genomic matrices using non-overlapping 400×400 sliding windows (2 Mb at 5 kb resolution) with a stride of 50 bins (250 kb). A block was deemed valid *only if* there were absolutely no zero-sum rows (i.e., missing data bins) in both the input Hi-C and the target Micro-C sub-matrices. This guarantees that metrics exclusively reflect the model's structural enhancement capabilities on high-quality sequencing regions.
2. **Clipping and Scaling:** To stabilize variance, all valid sub-matrices were clamped to a maximum value of 0.1. The SSIM was computed directly on these 2D blocks with the dynamic data range explicitly set to 0.1.
3. **Diagonal Exclusion for Correlations:** For PCC and SRCC calculations, the 2D blocks were flattened into 1D arrays. Crucially, we extracted only the upper triangular indices strictly above the main diagonal ($k = 1$). By explicitly excluding the main diagonal (distance $= 0$), we prevented the dominant extreme values of locus self-interactions from artificially inflating the correlation coefficients.

The final reported SSIM, PCC, and SRCC values represent the macroscopic average across all strictly validated genomic blocks for each evaluated chromosome.

Stratum-adjusted Correlation Coefficient (SCC) and Global Fidelity. Moving beyond localized pixel similarity, we employed the SCC (derived from the HiCRep algorithm) to comprehensively assess the global continuous distribution and distance-decay correlation. To prevent the inflation of correlation scores by background zeros or sequencing blind spots, we implemented a rigorous joint-masking strategy. A 1D genomic bin is deemed "valid" only if its marginal sum is strictly greater than zero in **both** the input Hi-C and the target Micro-C matrices. The final 2D evaluation mask is the outer product of these joint valid bins.

The specific computation process follows three steps:

1. **Clipping and Smoothing:** Extreme outlier values inherent in the heavy-tailed distribution of contact frequencies were clipped at a threshold of 0.1. A 2D uniform mean filter with a radius of $h = 1$ (i.e., a 3×3 sliding window) was then applied to mitigate local sparsity noise.
2. **Stratified Correlation:** The matrices were divided into continuous strata based on the genomic distance d. For each valid diagonal d, the Pearson correlation coefficient ρ_d between the smoothed predicted sequence (P_d) and the target sequence (T_d) was computed.
3. **Weighted Aggregation:** A weight W_d was assigned to each stratum:

$$W_d = N_d \cdot \sigma(P_d) \cdot \sigma(T_d)$$

   where N_d is the number of valid bins at distance d, and σ denotes the standard deviation.

The final unified SCC score is the weighted average of the correlations across all distance strata:

$$\text{SCC} = \frac{\sum_{d=0}^{k_{\max}} W_d \cdot \rho_d}{\sum_{d=0}^{k_{\max}} W_d}$$

Chromatin Loop Discovery and Matching Tolerance. Shifting from continuous correlations to discrete biological structures, chromatin loops represent high-frequency, focal interactions between distal genomic loci. We utilized `mustache` to detect loops on both the target Micro-C matrices and the model-predicted matrices. To compute the Precision, Recall, and F_1 scores, we mapped the genomic coordinates of all loop anchors to a 2D integer grid based on the 5 kb resolution.

Due to the inherent stochasticity of chromatin folding and matrix enhancement, forcing an exact pixel-to-pixel match is biologically overly restrictive. Therefore, we formulated the loop matching process as a 2D spatial nearest-neighbor search. Let L_{true} be the set of true loops and L_{pred} be the set of predicted loops. A predicted loop at 2D coordinate $(x_p, y_p) \in L_{\text{pred}}$ is classified as a True Positive (TP_{pred}) if there exists at least one true loop $(x_t, y_t) \in L_{\text{true}}$ such that the Euclidean distance between their anchors is within a defined tolerance radius τ:

$$\sqrt{(x_p - x_t)^2 + (y_p - y_t)^2} \leq \tau$$

To ensure computational efficiency over millions of genomic bins, this spatial querying was optimized using a K-Dimensional Tree (`cKDTree`) data structure. In our main text evaluations, we reported the results under a **1-bin tolerance** ($\tau = 1$), meaning the predicted loop anchor must fall within a strict 5 kb radius of the genuine Micro-C anchor. The corresponding metrics are formally defined as:

$$\text{Precision} = \frac{|TP_{\text{pred}}|}{|L_{\text{pred}}|}$$

$$\text{Recall} = \frac{|TP_{\text{true}}|}{|L_{\text{true}}|}$$

$$F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

where $|TP_{\text{true}}|$ is calculated symmetrically by querying the predicted loops against the true loops.

Topologically Associating Domain (TAD) Boundary Detection. Beyond focal loops, we evaluated the model's capacity to reconstruct macro-scale chromatin domains. TAD boundaries were identified by calculating the Insulation Score (IS) using the standard `cooltools.insulation` algorithm at 5 kb resolution. For a contact matrix M, the insulation score I_i at bin i is defined as the mean contact frequency within a sliding diamond window of size w centered at i:

$$I_i = \frac{1}{|W_i|} \sum_{(a,b)\in W_i} M_{a,b}, \quad W_i = \{(a,b) : |a-i| \le w \text{ and } |b-i| \le w\} \tag{9}$$

TAD boundaries are identified as local minima of the insulation score profile, with stronger boundaries corresponding to deeper local minima. IS was computed across multiple sliding window sizes, with a primary focus on a $w = 50$ kb window (10 bins at 5 kb resolution), the standard scale for mammalian TAD boundary analysis.

Similar to the loop evaluation, boundary matching was treated as a spatial nearest-neighbor search, optimized via a 1D `cKDTree` over the genomic indices. Let B_{true} and B_{pred} represent the sets of true and predicted 1D TAD boundary coordinates. A predicted boundary $x_p \in B_{\text{pred}}$ is considered a True Positive if there is a corresponding true boundary $x_t \in B_{\text{true}}$ satisfying $|x_p - x_t| \le \tau$.

Crucially, while loop detection utilized a 1-bin relaxation due to 2D sparsity, we applied an **absolutely strict 0-bin tolerance** ($\tau = 0$) for TAD boundary matching. This stringent criterion demands that the model-predicted insulation minimum aligns exactly within the identical 5 kb genomic bin as the high-resolution target. The Precision, Recall, and F_1 scores for TAD boundary recovery were calculated using the exact same harmonic mean formulation as described for the loop metrics. This zero-tolerance evaluation rigorously proves the model's sub-domain spatial accuracy.

C Qualitative visual assessment of chromatin architecture

To complement our quantitative metrics, we conducted a qualitative visual inspection of the generated contact maps. Figure 2 presents a representative genomic locus alongside its zoomed-in view, comparing the structural fidelity of various generative approaches against the low-resolution Hi-C input and the ground-truth Micro-C.

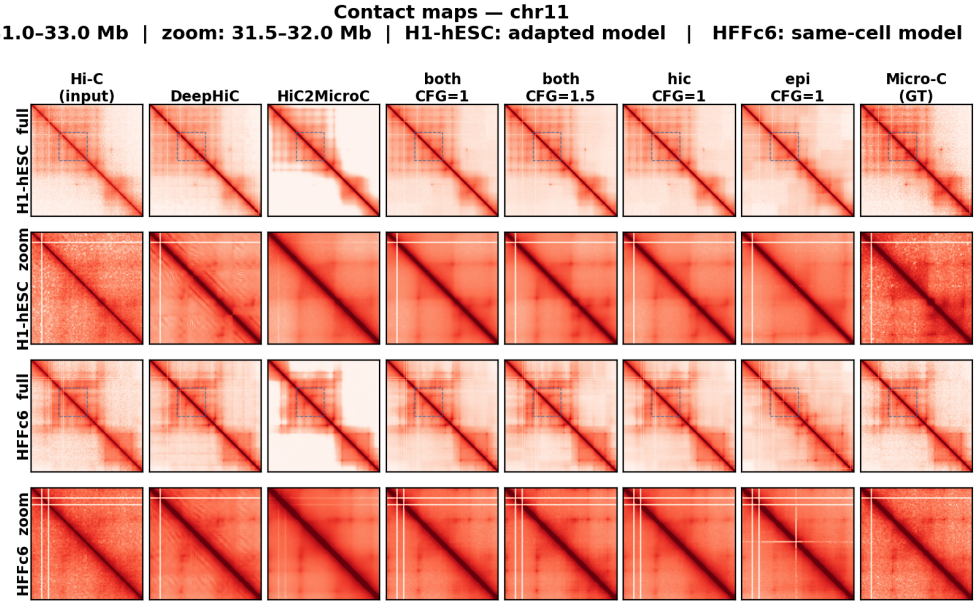

Figure 2: **Visual Comparison of Generated Contact Maps.** Qualitative evaluation of the target locus at both macroscopic (full) and microscopic (zoomed) scales. Our dual-conditioned model (`both`) faithfully reconstructs both global TADs and local fine-scale loops without the artificial oversmoothing or striation artifacts observed in the baseline methods.

Limitations of Existing Baselines. Visualizing the outputs of the baseline methods reveals distinct modality-specific artifacts. While **HiC2MicroC** successfully enhances certain focal interactions, it suffers from excessive smoothing. It uniformly suppresses both background contact frequencies and stochastic biological variation, resulting in an unnaturally stark contrast that fails to capture the realistic distance-dependent decay intrinsic to high-resolution Micro-C maps. Conversely, **DeepHiC** demonstrates a reasonable capacity to reconstruct broader global topology (e.g., macroscopic TADs). However, magnifying the contact map reveals conspicuous structural artifacts: irregular diagonal streaks (not necessarily parallel to the main diagonal), vertical banding patterns, and moiré-like synthetic textures that do not exist in the ground truth.

High-Fidelity Reconstruction of Our Framework. In stark contrast, our proposed framework—particularly in the `both` and `hic` configurations—exhibits remarkable visual concordance with the Micro-C ground truth. The model effectively recovers the crisp boundaries of Topologically Associating Domains (TADs) while accurately imputing fine-grained punctate loop patterns in the zoomed-in views. Importantly, it achieves this high-resolution enhancement while preserving the natural textural variance of the chromatin background, avoiding the synthetic artifacts that plague previous baselines.

D Extended loop and APA analysis

D.1 Chromatin loop discovery across tolerances

We performed genome-wide loop detection using `mustache` on held-out test chromosomes (chr8, 9, 11, 22). To account for the stochastic nature of chromatin folding, structural fidelity was evaluated across varying spatial tolerance thresholds ($\tau \in \{0, 1, 2, 3\}$ bins at 5 kb resolution). Complete Precision, Recall, and F_1 scores are summarized in Figure 3.

Figure 3 presents the full tolerance sweep across six panels. Panels (a) and (b) show same-cell evaluations on H1-hESC and HFFc6, respectively. At the standard 1-bin tolerance ($\tau = 1$, 5 kb), our `both` CFG=1 model achieves the highest F_1 on both cell types (H1-hESC: 0.4037; HFFc6: 0.5202), substantially ahead of HiC2MicroC (0.2685 and 0.4498) and DeepHiC (0.2034 and 0.3669). Notably, the performance advantage is sustained across all tolerance thresholds ($\tau = 0$ to $\tau = 3$), demonstrating that the improvement stems from genuine structural recovery rather than tolerance-level calibration. Moreover, the ungated ablation (`both` CFG=1.5, no gate: $F_1 = 0.3610$ on H1-hESC) consistently trails the gated variant ($F_1 = 0.3823$), confirming the spatial gate's contribution to fine-scale loop detection across all settings.

Panels (c) and (d) evaluate cross-cell generalization by comparing natively trained vs. transferred weights on the same target cell type. On H1-hESC, the natively trained `both` model achieves $F_1 = 0.4037$, while the adapted counterpart reaches 0.3517 (a 5.2 percentage-point drop), directly visualizing the remaining cross-cell batch effect. The `hic`-only configuration exhibits a smaller gap (0.3281 vs. 0.3092, a 1.9-point drop). The `epi`-only configuration is almost fully recovered after adaptation (0.1209 vs. 0.1196), in line with the improved cross-cell transferability of epigenomic signatures following adaptation.

Panels (e) and (f) decompose Precision and Recall across tolerances, revealing that our model's F_1 advantage derives primarily from substantially higher recall while maintaining competitive precision—a direct consequence of the deterministic coordinate anchoring mechanism described in Section 3.3.

D.2 Aggregate peak analysis (APA)

Aggregate Peak Analysis extracts and averages 21×21 pixel sub-matrices from the real Micro-C contact map centered at predicted loop coordinates, providing direct validation of biological signal quality. We evaluate True Positive (TP) and False Positive (FP) loop sets independently. Results are summarized in Figure 4.

The ground-truth (GT) Micro-C loops establish a baseline enrichment of $1.97\times$ on H1-hESC and $2.48\times$ on HFFc6. Low-resolution Hi-C is highly conservative: it recovers few loops (515 TP, 16% of GT on H1-hESC) but with exceptional enrichment ($2.77\times$, $1.41\times$ GT). Our framework (`both`,

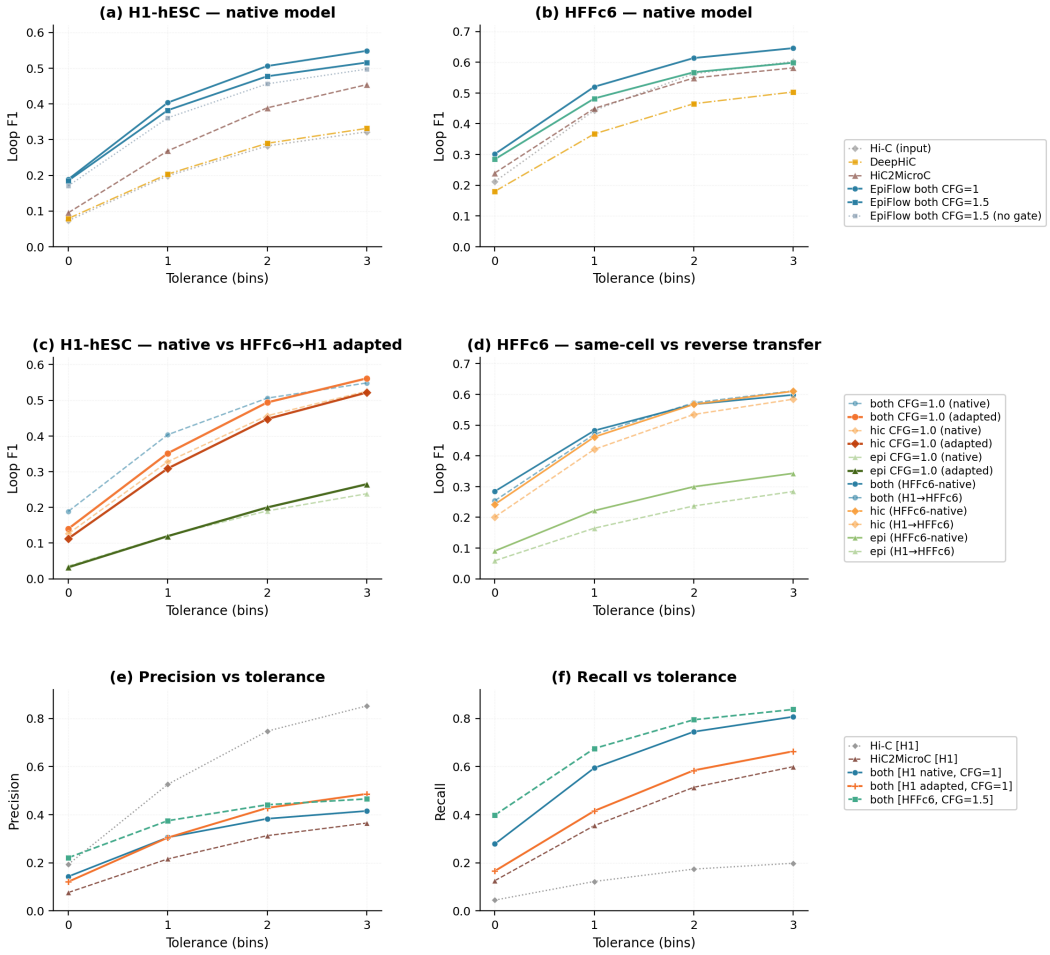

Figure 3: **Quantitative Evaluation of Chromatin Loop Recovery.** Loop detection F_1 scores as a function of varying spatial tolerance thresholds (τ). Results are presented for both the intra-cell (HFFc6) and cross-cell (H1-hESC, after adaptation) settings, including same-cell vs. cross-cell comparisons for key model configurations.

`CFG=1`) optimally balances this trade-off: 1,629 TP loops (51% of GT) at $2.14\times$ enrichment ($1.08\times$ GT) on H1-hESC, and 3,478 TP (71% of GT) at $2.62\times$ on HFFc6.

Critically, FP predictions are not random noise—they exhibit enrichment above 1.0 ($1.30\times$ on H1-hESC, $1.39\times$ on HFFc6), confirming genuine sub-threshold contact enrichment. Our model maintains strong TP/FP discrimination (ratio $1.64\times$ on H1, $1.88\times$ on HFFc6).

The epigenomic-only configuration (`epi, CFG=1`) shows how adaptation changes cross-cell behavior: on HFFc6 it identifies 1,511 TP loops at $2.77\times$ enrichment, while on H1-hESC, after adaptation, it identifies 544 TP loops at $2.24\times$ enrichment (FP enrichment $1.06\times$); the remaining gap to the dual-modality model confirms that epigenomic features remain less informative without the Hi-C backbone.

E Structural consistency analysis

We present three complementary analyses that collectively validate the structural consistency of our generative framework: distance-decay P(s) curves, per-stratum SCC decomposition, and the learned Toeplitz prototype.

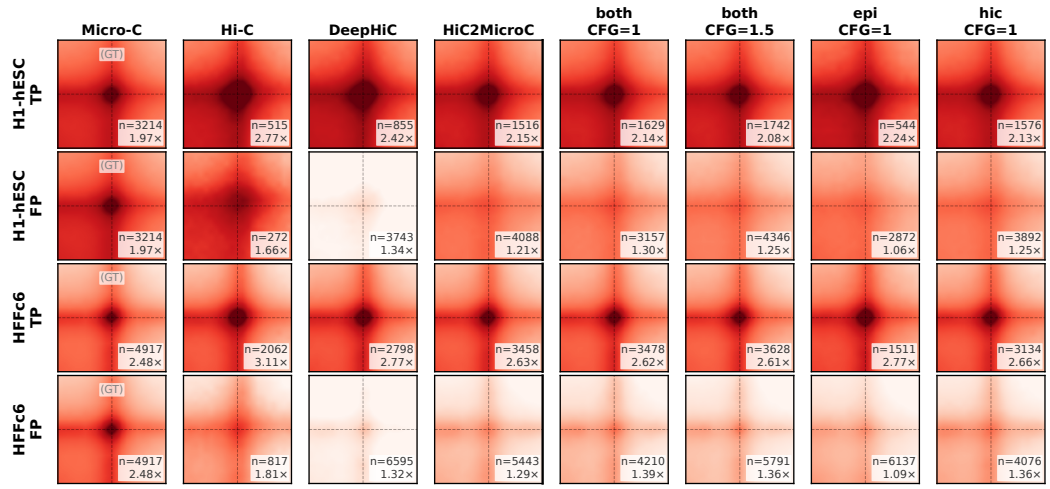

Figure 4: **Aggregate Peak Analysis (APA) of Predicted Chromatin Loops.** Aggregated Micro-C signal at TP and FP loop predictions on H1-hESC and HFFc6. Our dual-conditioned model (`both, CFG=1`) achieves strong focal enrichment for TPs while maintaining clear discrimination against FPs.

E.1 Distance decay P(s) analysis

To quantify how faithfully each method preserves the characteristic distance-decay profile of chromatin contacts, we compute P(s) curves and compare them directly with the Micro-C ground truth.

Figure 5 presents the analysis over 0–2 Mb. On H1-hESC, the raw Hi-C achieves $r = 0.9975$, our model preserves $r = 0.9948$, while DeepHiC ($r = 0.9760$) and HiC2MicroC ($r = 0.9259$) show progressively larger deviations. The ranking is consistent on HFFc6. Baseline deviations concentrate at intermediate distances (0.5–2 Mb), directly confirming the oversmoothing hypothesis from the main text.

E.2 Per-stratum SCC decomposition

To dissect the SCC behavior, we decompose the unified score into per-stratum correlations ρ_d as a function of genomic distance d.

The per-stratum analysis reveals distinct distance-dependent behaviors. DeepHiC maintains high ρ_d across distances ($\bar{\rho}_d \approx 0.75$), consistent with global oversmoothing. HiC2MicroC matches or exceeds DeepHiC at short distances ($\rho_d \approx 0.85$ at distances ≤ 0.5 Mb) but collapses at long distances ($\rho_d \approx 0.05$ on H1-hESC at > 1.5 Mb). Our method (`both` CFG=1) achieves the highest short-range correlations among the generative models ($\bar{\rho}_d \approx 0.87$ at distances ≤ 0.5 Mb) while maintaining $\rho_d \approx 0.37$ at > 1.5 Mb (0.49 with CFG=1.5). The epi-only configuration on H1-hESC reaches $\bar{\rho}_d \approx 0.26$ after adaptation (about 0.04 at > 1.5 Mb), marking a clear improvement in overall tracking while remaining the weakest at short range.

E.3 Learned Toeplitz prototype

The Toeplitz prototype matrix $T \in \mathbb{R}^{H \times W}$, parameterized by a learnable 1D distance-bias vector $w \in \mathbb{R}^L$ with $T_{i,j} = w_{|i-j|}$, serves as the distance-decay baseline for unconditioned CFG inference. We extract w from both HFFc6-trained and H1-hESC-trained models. The learned w vectors are nearly identical across cell types (Pearson $r = 0.989$), confirming that the prototype captures a cell-type-invariant structural property rather than cell-specific artifacts.

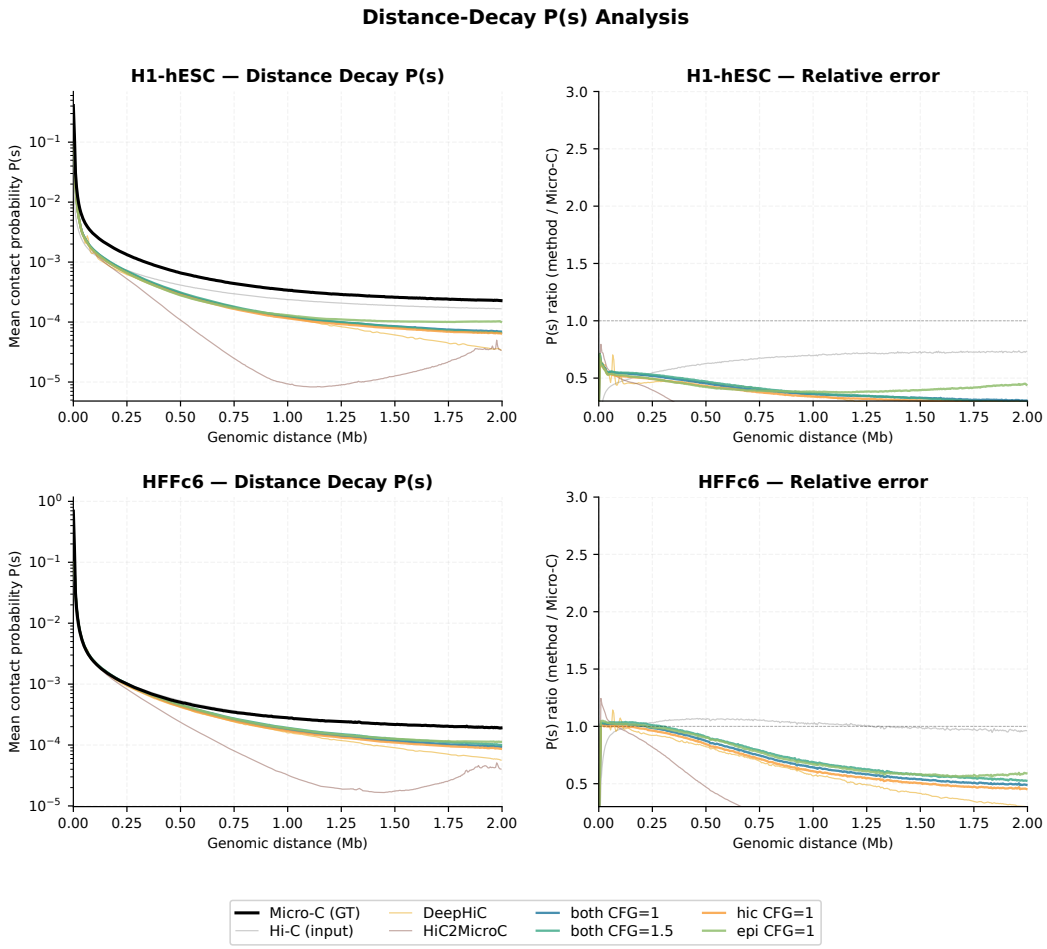

Figure 5: **Distance-decay ($\mathbf{P}(s)$) analysis.** Top row: full $\mathrm{P}(s)$ curves (left) and per-distance relative error vs. Micro-C (right) for H1-hESC. Bottom row: same for HFFc6. Our method (`both` CFG=1) closely tracks the Micro-C decay profile (Pearson $r = 0.9948$ on H1, 0.9813 on HFFc6), while baseline methods—particularly DeepHiC ($r = 0.9760$) and HiC2MicroC ($r = 0.9259$)—systematically deviate, confirming the oversmoothing hypothesis.

Figure 7 displays the full 600×600 Toeplitz matrix. The smooth, radially symmetric decay pattern confirms that the prototype learns a physically meaningful baseline: contact probability decreases monotonically with genomic distance, matching the characteristic behavior of real chromatin contact data.

F Broader societal impacts

As foundational research in computational biology, our continuous-time Flow Matching framework aims to advance the fundamental understanding of 3D genome architecture. While not directly deployed in clinical settings, this work carries practical implications for the scientific community.

Positive Impacts. By enabling the reliable generation of ultra-high-resolution Micro-C landscapes from widely available, standard-depth Hi-C data, our framework democratizes access to fine-scale genomic structural analysis. This significantly reduces the prohibitive sequencing costs associated with mapping 3D genomes, allowing research laboratories with limited funding to computationally explore complex enhancer-promoter interactions and accelerate discoveries in basic molecular biology.

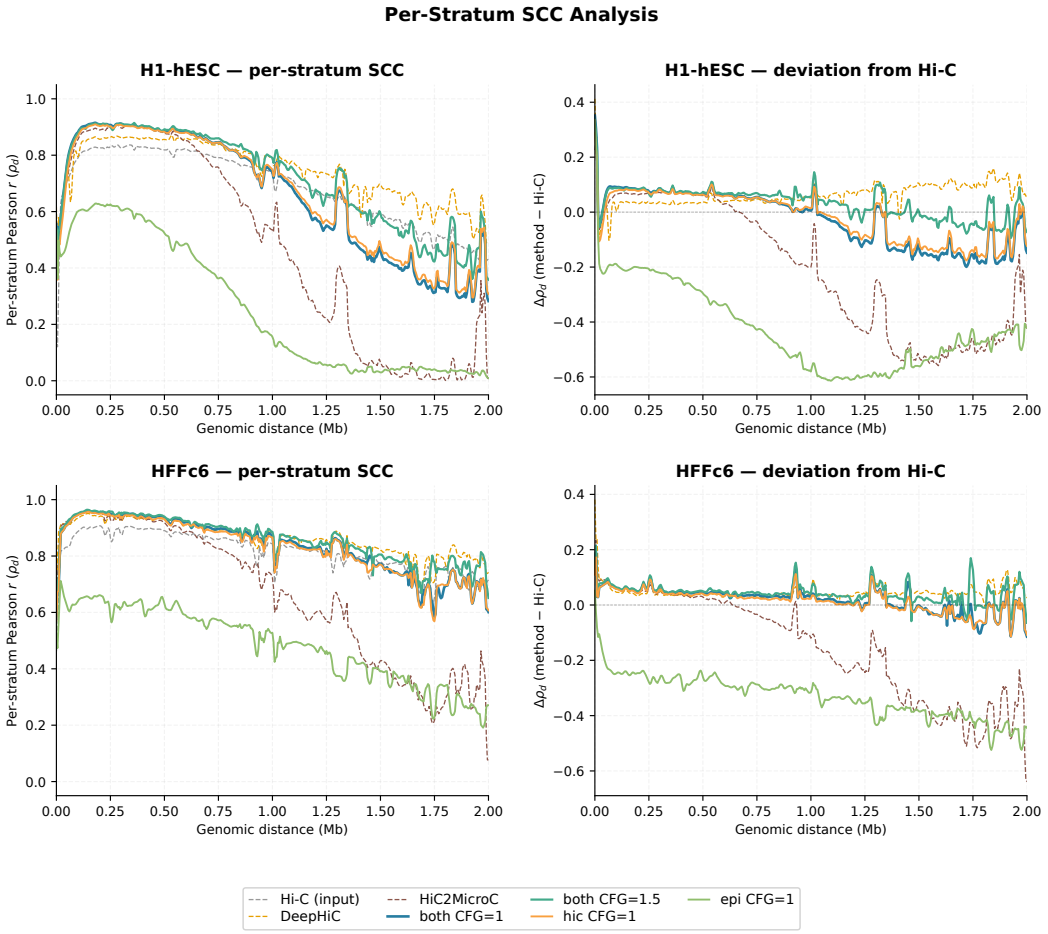

Figure 6: **Per-stratum SCC decomposition.** Left column: per-stratum Pearson correlation ρ_d between each method and Micro-C, as a function of genomic distance. Right column: deviation $\Delta\rho_d$ relative to the Hi-C input baseline. DeepHiC maintains high ρ_d across distances, while our method (`both` CFG=1) stays closer to the Hi-C reference profile and achieves the highest short-range correlations among generative models.

Potential Negative Impacts and Mitigation. Despite these advancements, we acknowledge a critical downstream risk when deploying generative AI in fundamental biology: **Scientific Misdirection via "Hallucinations".** Like all generative models, our framework carries the inherent risk of producing false-positive structural features (i.e., "hallucinated" chromatin loops or TAD boundaries). In the field of genomics, downstream validations of these structures often require highly expensive and time-consuming wet-lab experiments, such as CRISPR-Cas9 genome editing or 3D DNA FISH. If biologists blindly over-rely on computationally hallucinated interactions to design these experiments, it could lead to severe scientific misdirection and the misallocation of valuable research funding.

To mitigate this risk, our framework explicitly enforces distance-dependent structural constraints (via the symmetric Toeplitz prototype) to minimize physically implausible structures. Furthermore, we strongly advocate that AI-generated chromosomal contact maps should be utilized strictly as *hypothesis-generation tools* to guide research, rather than as definitive replacements for ground-truth sequencing data.

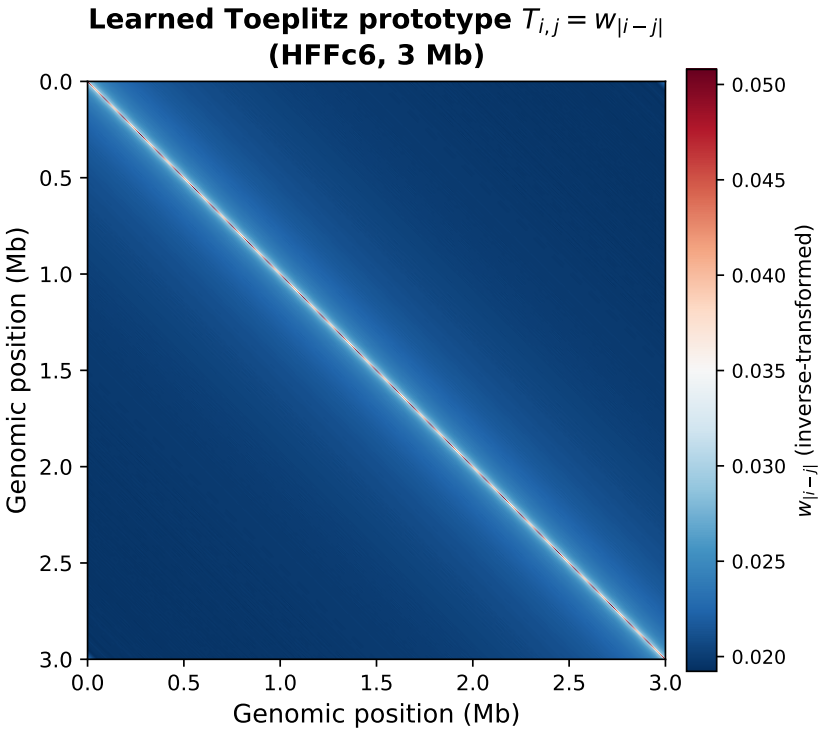

Figure 7: **Learned Toeplitz distance-decay prototype.** The 2D Toeplitz matrix $T_{i,j} = w_{|i-j|}$ for the HFFc6 model, visualized over the full 3 Mb genomic window (600 bins $\times$ 5 kb) after inverse transformation from the model's normalized space. The diagonal-dominant structure with smooth distance-dependent decay directly encodes the physical prior into the CFG mechanism.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The contributions regarding our high-resolution Micro-C enhancement model and its cross-cell line generalization are explicitly stated in the Section 1 and are supported by experimental results in Section 4.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Limitations are discussed in Section 5: weaker direct cross-cell epigenomic transfer that is recovered by adaptation (Epi-only F1 drops to 0.023 under direct transfer; retraining on the target cell type recovers F1=0.404, and an indirect adaptation that fine-tunes only the conditioning modules restores epigenome-only transfer to near-native performance (0.120 vs. 0.121) while raising both-modality F1 to 0.352), occasional false-positive loops under extreme epigenetic signal density, and the need for broader validation across cell lines and species. Broader societal risks are addressed in Appendix F.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best

judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [N/A]

   Justification: This work is primarily empirical and focuses on the application of generative modeling to 3D genomics; it does not introduce new theoretical theorems that require formal proofs.

   Guidelines:

   - The answer [N/A] means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: All experimental details are provided: dataset accession IDs (Appendix Table 4), preprocessing pipeline (Section A.2), full architecture specification (Section 3), optimizer and hyperparameters (Section A.3), and evaluation protocols (Section 4). The companion code package includes scripts to reproduce all major results.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.

(c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: All 14 datasets are publicly available from the 4DN Data Portal and ENCODE Portal with accession IDs listed in Appendix Table 4. The complete codebase is released with all model definitions, training/inference/evaluation scripts. Dependencies are documented in the accompanying README.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: All critical experimental settings, including optimization parameters, data splitting, dynamic cropping strategies, and the rationale behind hyperparameter choices are explicitly detailed in Appendix A.3.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [No]

Justification: Error bars from multiple random-seed runs are not reported due to computational cost (14 GPU-hours per training run). However, all metrics aggregate millions of genomic bin pairs across four full-length chromosomes, and results are consistent across both cell types and CFG scales. The cross-cell zero-shot setting provides a stricter validation than repeated same-cell runs.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Detailed hardware specifications and execution times are thoroughly documented in Appendix A.4.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: We have read the NeurIPS Code of Ethics and confirm that our research conforms to it. All genomic data used in this study (H1-hESC and HFFc6) are obtained from public repositories (e.g., 4DN Data Portal) and used in accordance with their established ethical and privacy guidelines.

Guidelines:

- The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: A comprehensive discussion of broader impacts is provided in Appendix F. We highlight the positive impact of democratizing high-resolution genomic analysis by reducing sequencing costs, while critically addressing the negative risk of generative "hallucinations" causing misdirection and wasted resources in downstream wet-lab validations.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The proposed model is designed specifically for 3D genomic data enhancement and does not pose high risks for societal misuse like large language models or generic image generators.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.

- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: All datasets, baseline code, and software libraries are properly credited. 4DN and ENCODE data use policies are respected; accession IDs are listed in Appendix Table 4. Baseline methods (DeepHiC, HiC2MicroC) are cited with their original publications. All software dependencies (PyTorch, cooler, cooltools, torchcfm, etc.) are open-source and cited.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: The released code includes model checkpoints, all training/inference/evaluation scripts, and a README with installation and usage instructions. Architecture, hyperparameters, and data preprocessing are fully documented in the paper appendices.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

Justification: This research does not involve crowdsourcing or active human participants; it utilizes publicly available genomic datasets that do not require human subject experimentation protocols.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [N/A]

Justification: This work utilizes only de-identified, publicly available genomic datasets from the 4DN Data Portal and ENCODE, which do not qualify as human subjects research requiring IRB approval for the secondary analysis performed in this study.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: LLMs were only used for language editing and formatting of the manuscript, and were not part of the core methodology.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.