bioRxiv preprint doi: https://doi.org/10.1101/2025.11.18.689171; this version posted November 19, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Evo2HiC: a multimodal foundation model for integrative analysis of genome sequence and architecture

Tangqi Fang*[1], Xiao Wang*[1,2], Zhiping Xiao[1], Shengqi Hang[1], Ghulam Murtaza[2], Junwei Yang[1], Hanwen Xu[1], Anupama Jha[2], William Noble†[1,2], and Sheng Wang†[1]

[1]Paul G. Allen School of Computer Science and Engineering, University of Washington, Seattle, WA, USA.
[2]Department of Genome Sciences, University of Washington, Seattle, WA, USA

Abstract

Understanding how genomic sequences shape three-dimensional (3D) genome architecture is fundamental to interpreting diverse biological processes. Although previous studies have shown that sequence information can predict 3D genome architecture, they fall short in capturing cell type–specific structures because they are trained solely on sequence inputs. The widely available Hi-C data, which contain rich structural information across biosamples, can provide complementary features to sequence data for studying cell type–specific architectures. Recently, DNA foundation models have demonstrated encouraging performance in capturing long-range genomic dependencies, holding promise for modeling chromatin interactions. However, the extremely high computational cost of running these models limits their applicability to Hi-C analysis, which requires genome-wide sequence embeddings. Here, we present Evo2HiC, a multimodal foundation model that jointly models genomic sequences and structures to study cell type-specific chromatin structure. The key idea of Evo2HiC is to distill a large-scale DNA foundation model, Evo 2 (7B), into a compact encoder, while guiding the distillation with Hi-C data to preserve genomic features critical for 3D genome analysis. The model supports two types of encoders, one that operates directly on DNA sequences, and a second that additionally takes as input corresponding Hi-C data. Using the DNA-only encoder and predicting Hi-C contact matrices, Evo2HiC improved Spearman correlation by 10.9% over Orca. Moreover, by jointly embedding Hi-C and sequence information Evo2HiC achieved the best overall Pearson correlation when predicting five representative epigenomic assays. Interpretation analysis of Evo2HiC revealed its ability to identify cell type–specific sequence motifs that explain changes in epigenomic signals. Finally, we demonstrated the cross-species generalizability of Evo2HiC on 177 species from the DNA Zoo dataset for Hi-C resolution enhancement. In summary, Evo2HiC is a foundation model that integrates genome sequences and 3D chromatin structure information, substantially reduces computational cost while maintaining state-of-the-art accuracy on predicting various epigenomic signals and genome architecture, enables the identification of cell type-specific motifs, and demonstrates robust generalizability across species.

*These authors contributed equally
†Correspondence: william-noble@uw.edu, swang@cs.washington.edu

bioRxiv preprint doi: https://doi.org/10.1101/2025.11.18.689171; this version posted November 19, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

Understanding how genetic information flows from the DNA sequence to downstream biological processes is fundamental across diverse biological contexts, including development [1, 2, 3], evolution [4, 5, 6], and disease [7, 8, 9, 10, 11, 12]. Within the nucleus, this genetic information is organized into a three-dimensional structure that profoundly influences essential genome functions, including DNA replication, transcription, cell differentiation, and cellular senescence [13]. To characterize this spatial organization, a variety of experimental techniques, such as Hi-C [14], GAM [15], SPRITE [16], and HiChIP [17], have been developed to map chromatin architecture and DNA–DNA interactions across the genome, generating a large number of contact matrices in which the axes represent genomic loci and the matrix values reflect their spatial proximity in three dimensions.

Meanwhile, genomic sequence encodes features such as CTCF binding motifs and CpG islands that act as key determinants of chromatin loops, domains, and compartments, thereby shaping the three-dimensional genome architecture [18, 19]. Building on this principle, several computational methods, including Orca [20], Akita [21], and DeepC [22], have been developed to predict chromatin organization directly from DNA sequences, revealing the underlying relationships between chromatin sequences and structure.

However, because current approaches are trained solely using sequences as inputs, they need to be retrained when analyzing a new cell type. This failure to generalization across cell types makes it challenging for these models to characterize cell type-specific chromatin interaction patterns. One potential solution is to jointly train a model using genome sequences and Hi-C data. In contrast to genome sequences, which allow us to extract high-resolution but 1D features, a Hi-C contact matrix directly delivers 2D patterns but is typically analyzed at coarse resolution (e.g., 10 kb [14]) due to sequencing depth constraints. Therefore, high-resolution 1D features derived from genome sequences and 2D features from low-resolution contact matrices provide complementary information for deciphering cell type-specific chromatin structures.

Jointly analyzing Hi-C data and genome sequences requires a model to process millions of bases as input, posing substantial challenges for capturing long-range dependencies. Recently, a series of DNA sequence foundation models have demonstrated strong performance across diverse genomic tasks [23, 24, 25, 26] due to their ability to model long-range dependencies. For instance, the recently developed DNA foundation model Evo 2 can process sequence windows up to one million base pairs [27], demonstrating its capacity to capture distal genomic interactions. Despite this potential, applying DNA foundation models to chromatin structure prediction remains computationally demanding for several reasons. First, Hi-C data represent pairwise genomic interactions, requiring simultaneous analysis of two DNA sequences, which doubles computational cost. Second, embeddings for the whole genome must be computed even when only near-diagonal contacts are of interest. Third, models such as Evo 2 require high-end GPUs, such as NVIDIA H100s, to support efficient inference, further limiting accessibility for many experimental laboratories lacking such computational resources.

To this end, we propose Evo2HiC, a multimodal foundation model for jointly analyzing DNA sequence and Hi-C contact matrices. The key idea is to distill the powerful but expensive DNA foundation model Evo 2 into a much smaller model that maintains the long-range modeling ability of Evo 2 but requires fewer computational resources. Instead of directly distilling Evo 2 to a small model, we jointly project Evo 2 and our model to a shared Hi-C embedding space. This approach allows the distillation to focus on features that are critical to the chromatin structure. The resulting embedding space can be used as an encoder for downstream applications that involve Hi-C data and DNA sequence data. This distilled encoder reduces the time and memory requirements for downstream applications by a factor of 500 compared to Evo 2.

We demonstrate the versatility, effectiveness, and interpretability of Evo2HiC for studying cell type-specific chromatin architecture. We observe a strong consistency between Evo 2–based embeddings and those produced by Evo2HiC. Evo2HiC achieves the best performance in predicting Hi-C contact matrix directly from sequence, with a 10.9% improvement in Spearman correlation over Orca [28]. Moreover, Evo2HiC achieves state-of-the-art performance in predicting five representative types of epigenomic signals, including DNase, CTCF, H3K27ac, H3K27me3, and H3K4me3. Interpretation of the model further reveals cell type–specific sequence motifs that explain observed changes in epigenomic activity. The distilled encoder in Evo2HiC also exhibits strong cross-species generalizability when using only Evo 2 human embeddings. By jointly leveraging genomic sequences and low-coverage contact matrices, Evo2HiC achieves the best performance over four existing methods for Hi-C resolution enhancement across 177 species.

bioRxiv preprint doi: https://doi.org/10.1101/2025.11.18.689171; this version posted November 19, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

In summary, Evo2HiC is a multimodal foundation model that jointly models genome sequences and 3D chromatin structures. By distilling large-scale DNA language models into a compact architecture grounded in Hi-C data, Evo2HiC substantially reduces computational cost while maintaining high predictive accuracy across diverse tasks, including predicting the Hi-C contact matrix from DNA sequence, epigenomic signal prediction, and Hi-C resolution enhancement, and further demonstrates cell type-specific interpretability and cross-species generalizability.

Results

Overview of Evo2HiC

Evo2HiC aims to learn a lightweight encoder capable of capturing megabase-scale long-range interactions that are critical for Hi-C data analysis (**Fig. 1**). To achieve this, we distill the 7 billion-parameter Evo 2 model into a compact CNN-based genome sequence encoder consisting of seven layers and totaling 3.6 million parameters, trained on 1.2 million 2 kb human genomic bins. Given the nearly 2000-fold reduction in model size, it is infeasible for the CNN encoder to retain all information from Evo 2. Therefore, we encourage the model to focus on genomic features most relevant to the chromatin structure. To this end, we leverage Hi-C data to guide the distillation process. Specifically, we optimize two contrastive learning objectives: one aligns the sequence embeddings from the CNN encoder with those from Evo 2, and the other aligns the CNN-based embeddings with Hi-C patch embeddings produced by a Hi-C encoder. We deliberately avoid direct alignment between Evo 2 and the Hi-C encoder to mitigate the computational burden arising from large-scale 2D contact matrices and genome-wide embeddings. Following these two alignment steps, the resulting CNN-based genomic sequence encoder serves as our encoder for embedding genomic sequences. The co-embedded representation, which concatenates embeddings from both Hi-C contact maps and genomic sequences across the entire human genome, further enables downstream applications, such as Hi-C resolution enhancement and epigenomic signal prediction.

Evo2HiC preserves the genomic sequence features from Evo 2

We first examined whether the distilled CNN-based encoder can preserve genomic sequence features from Evo 2. To this end, we compared pairwise similarities between 2 kb human genomic sequences embedded by Evo 2 and by Evo2HiC (**Fig. 2a,b**). The embeddings showed a strong correlation of 0.63, markedly higher than the 0.13 correlation achieved by the same CNN architecture trained without Evo 2 distillation (see **Methods**). Given that Evo 2 was pretrained on 128,411 species, we next asked whether the distilled encoder trained only on human data could generalize to other species. Indeed, when evaluated on 78,255 2 kb mouse genomic sequences, Evo2HiC achieved substantially higher correlation with Evo 2 compared to the non-distilled CNN encoder (**Fig. 2c,d**), indicating successful transfer of cross-species genomic representations.

Having confirmed that our model retains Evo 2–derived genomic features, we next assessed whether these features benefit Hi-C analysis. On a held-out set of genomic sequence and Hi-C patch pairs, we compared the performance of identifying the corresponding Hi-C patch using sequence embeddings distilled from Evo 2 (**Fig. 2e-f, Supplementary Figure 1**). Our embeddings achieved higher recall and lower average rank than those from the encoder trained without Evo 2 on both human and mouse datasets, demonstrating the relevance of Evo 2 features for capturing chromatin organization. The performance gain was more pronounced in mouse, whose sequence data were never used during training, suggesting improved cross-species generalizability likely stemming from Evo 2's multi-species training.

Beyond accuracy, Evo2HiC offers substantial computational efficiency. In contrast to Evo 2's computationally expensive inference, Evo2HiC employs an efficient CNN-based architecture, achieving up to 500-fold faster inference at a genomic sequence length of 70 kb (**Fig. 2g, Supplementary Figure 2**).

Finally, we explored whether the co-embedding space learned from genomic sequences and Hi-C contact matrices could capture functional genomic activity. Visualization of the epigenomic scores from different assays (H3K27me3, H3K27ac, DNase, CTCF, H3K4me3) revealed clear spatial patterns in the co-embedding space (**Fig. 2h-j, Supplementary Figure 3**), highlighting its potential for predicting diverse epigenomic signals.

bioRxiv preprint doi: https://doi.org/10.1101/2025.11.18.689171; this version posted November 19, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

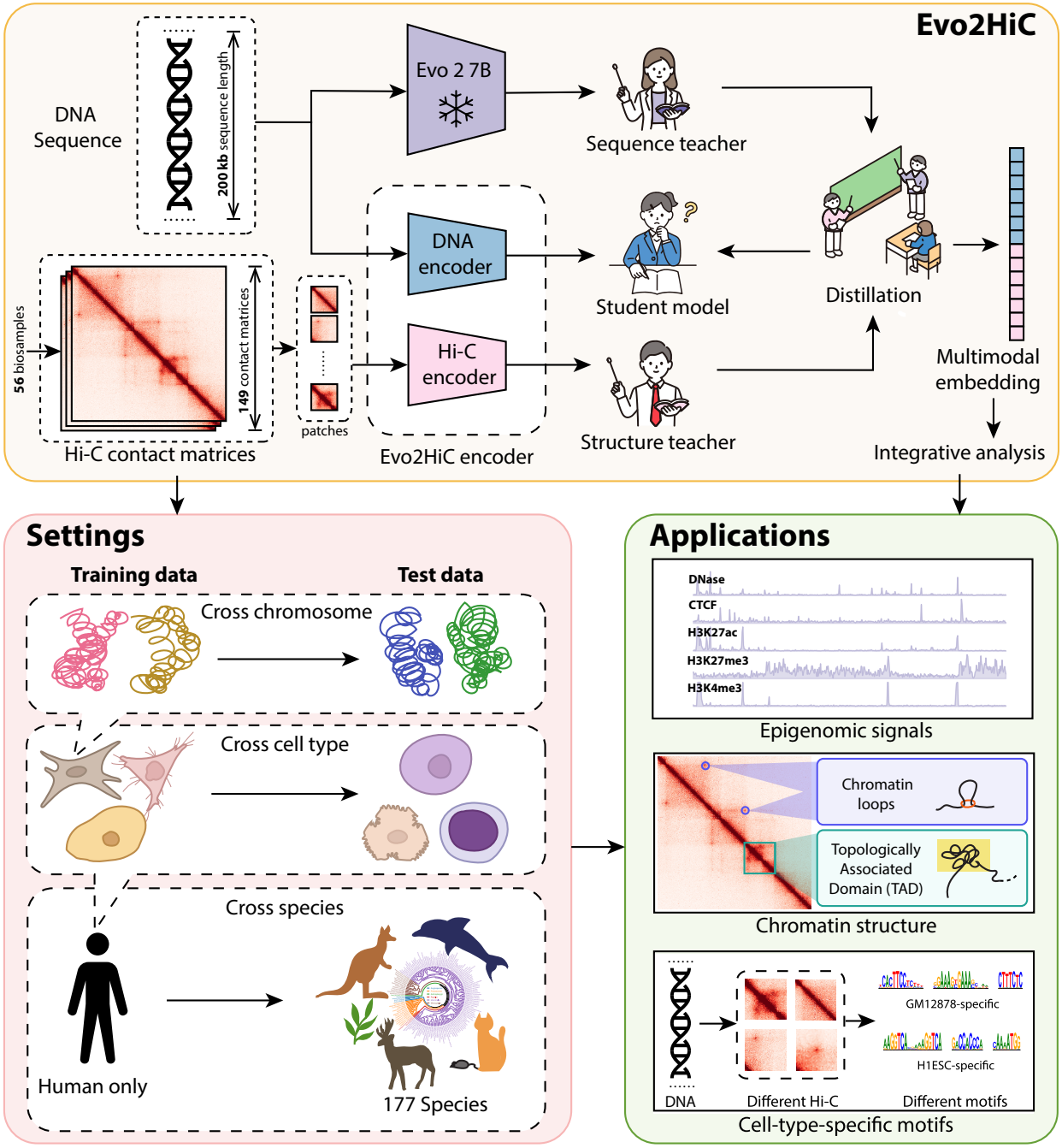


Figure 1: **Overview of Evo2HiC.** Evo2HiC is a multimodal foundation model that integrates Hi-C contact map and DNA sequence. It exploits distillation to learn a compact CNN-based sequence encoder from large-scale DNA foundation model Evo 2, while encouraging this distillation to focus on genomic features that are more critical to the chromatin structure. By integrating sequence and contact matrix, Evo2HiC allows cross cell-type and cross-species generalizability, further improving the performance on predicting epigenomic signals, chromatin structure analysis, and cell type-specific motif identification.

bioRxiv preprint doi: https://doi.org/10.1101/2025.11.18.689171; this version posted November 19, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

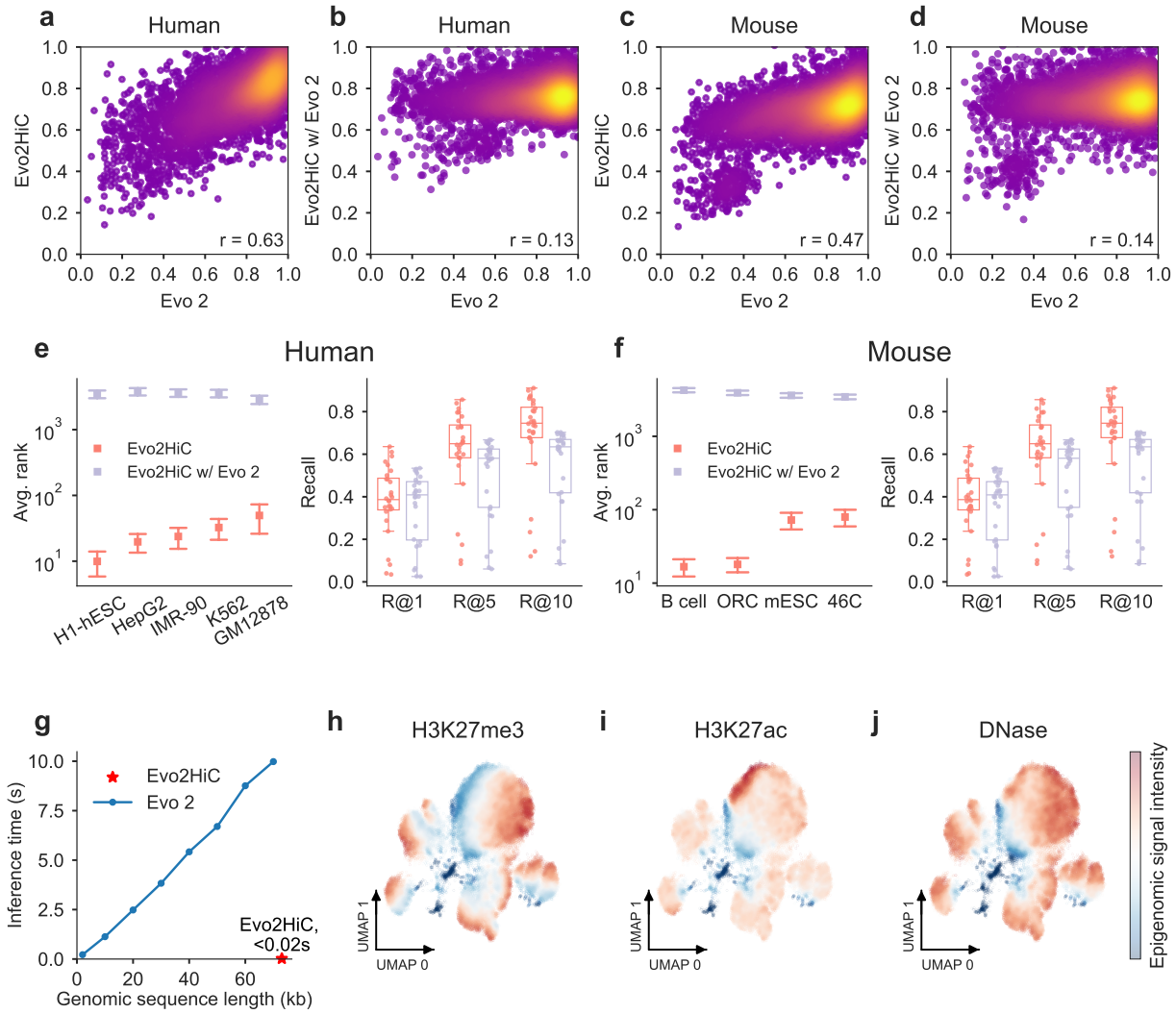


Figure 2: **Examining the embedding space of Evo2HiC. a,c,** Scatter plots comparing the embedding similarity between pairs of 2 kb sequences using Evo 2 and Evo2HiC on human (**a**) and mouse (**c**). **b,d,** Scatter plots comparing the embedding similarity between pairs of 2 kb sequences using Evo 2 and a CNN encoder trained without Evo 2 distillation on human (**b**) and mouse (**d**). **e,f,** Comparison between Evo2HiC and the Evo2HiC without Evo 2 on retrieving the most similar Hi-C patch using the corresponding genomic sequences in terms of average rank and recall on human (**e**) and on mouse (**f**). **g,** Comparison of inference cost for genomic sequences of different lengths between Evo 2 and Evo2HiC. The Evo 2 cost increases with larger genomic sequence length, whereas Evo2HiC (red star) has a low inference cost. **h-j,** UMAP visualization of the embeddings of genomic sequences by Evo2HiC, colored by epigenomic signals of H3K27me3 (**h**), H3K27ac (**i**), DNase (**j**).

bioRxiv preprint doi: https://doi.org/10.1101/2025.11.18.689171; this version posted November 19, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

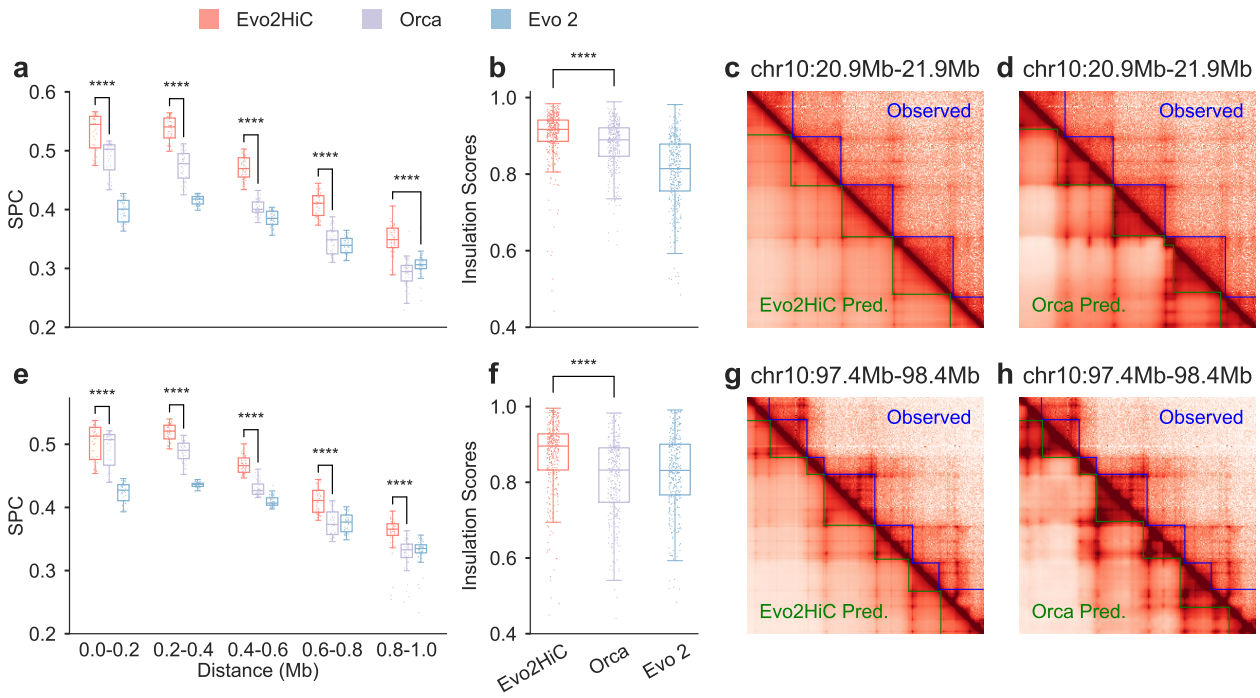


Figure 3: **Genome structure prediction from sequence. a–h,** Comparison among Evo2HiC, Orca, and Evo 2 for predicting Hi-C contact maps from genomic sequence input. Metrics are Spearman correlation on Hi-C contact counts **(a, e)** and Pearson correlation on insulation score **(b, f)**. **a–d** are results on the H1ESC cell line and **e–h** are results on the HFF cell line.

Evo2HiC predicts Hi-C contact matrix using genome sequences

Before jointly analyzing genome sequence and Hi-C data, we first evaluated the ability of the sequence encoder in Evo2HiC to predict Hi-C contact matrices (**Fig. 3a–h, Supplementary Figure 4,5,6**). Specifically, a U-Net–based decoder was trained to generate Hi-C contact patches from the learned sequence embeddings. We compared Evo2HiC with Orca, which employs a hierarchical sequence encoder to predict contact matrices, and with Evo 2, which uses the same U-Net–based decoder as our model. The latter comparison serves as an ablation to assess the intrinsic quality of our sequence encoder.

Across two human cell lines, H1ESC and HFF, Evo2HiC outperformed both Orca and Evo 2 in terms of Spearman correlation on the Hi-C contact counts (**Fig. 3a,e, Supplementary Figure 4**). When evaluated by insulation score (**Fig. 3b,f, Supplementary Figure 5**), which reflects the organization of topologically associating domains (TADs), Evo2HiC also achieved the highest performance, demonstrating its capacity to capture fine-grained 3D chromatin features. Notably, the improvement over Orca was even more pronounced for long-range contacts, partially due to Evo2HiC's ability to capture long-range interactions (**Fig. 3a,e**).

Evo2HiC predicts epigenomic profiles

Next, we leveraged the joint embedding of Hi-C and genomic sequences learned by Evo2HiC to predict epigenomic signals using a transformer-based decoder trained on five representative assays: DNase accessibility and ChIP-seq for CTCF, H3K27ac, H3K27me3, and H3K4me3. In the cross-chromosome evaluation, Evo2HiC substantially outperformed the Hi-C–only baseline and Evo 2, both of which used the same decoder as Evo2HiC but took only Hi-C data alone or Evo 2 embeddings as input, achieving an average improvement of 34.7% over the Hi-C–only baseline and 26.2% over Evo 2 across all assays (**Fig. 4a,b, Supplementary Figure 7a,b**). Under the more challenging cross-cell-line and cross-chromosome setting, Evo2HiC also surpassed both models in three out of five assays (**Fig. 4c, Supplementary Figure 7c**). A closer examination of the predicted profiles further revealed that Evo2HiC captures fine-grained genomic activity (**Fig. 4d**),

bioRxiv preprint doi: https://doi.org/10.1101/2025.11.18.689171; this version posted November 19, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

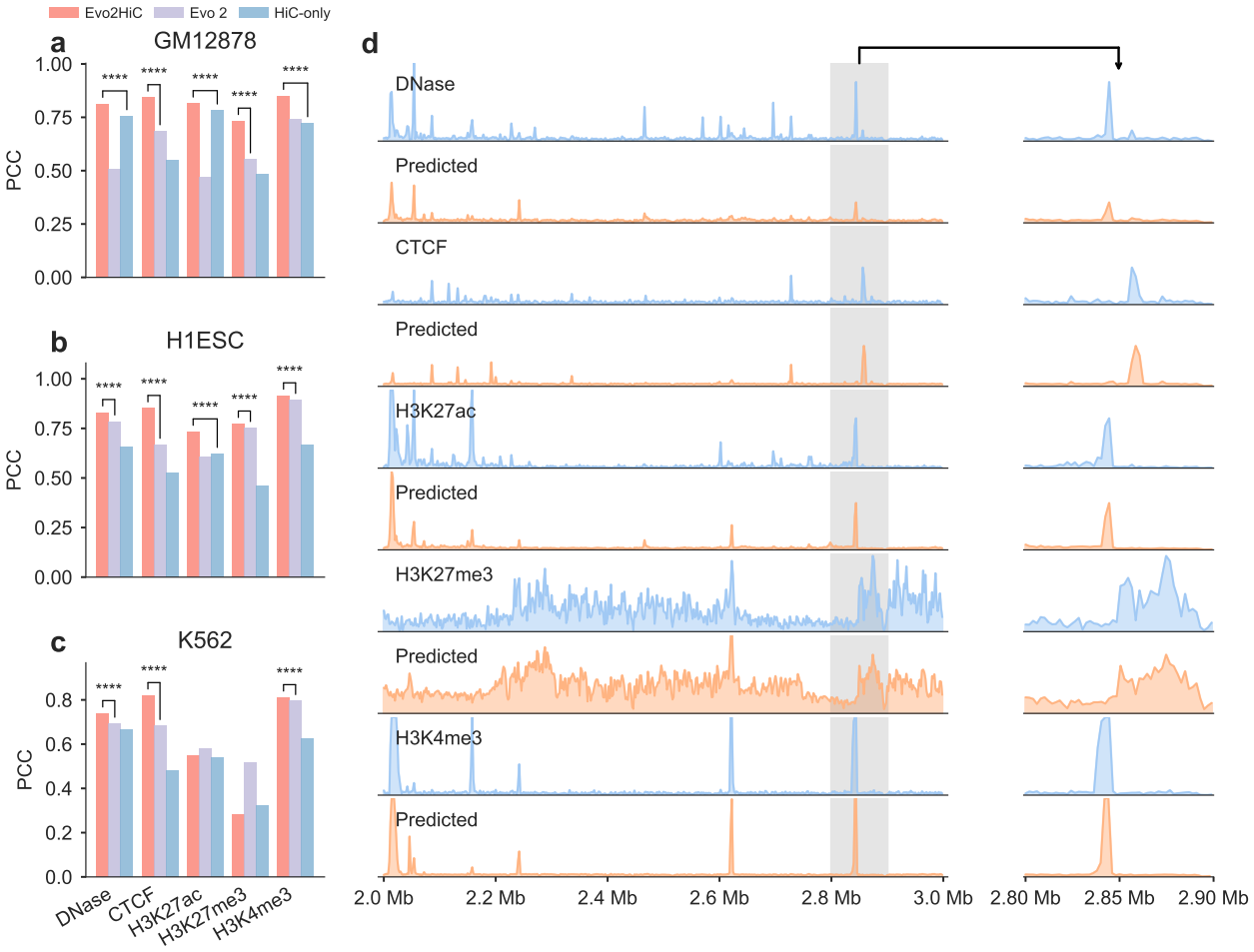


Figure 4: **Comparison on predicting epigenomic signals. a,b,** Comparison on predicting epigenomic signals in the cross chromosome setting on GM12878 (**a**) and H1ESC (**b**). **c,** Comparison on predicting epigenomic signals in the cross cell line and cross chromosome setting on K562. **d,** Comparison between the predicted epigenomic signal by Evo2HiC and the observed epigenomic signals in the GM12878 cell line (chr9:2 Mb - 3 Mb).

bioRxiv preprint doi: https://doi.org/10.1101/2025.11.18.689171; this version posted November 19, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

reflecting its strong ability to integrate sequence and structural representations.

Evo2HiC identifies interpretable cell type-specific motifs

Identifying motifs that drive epigenomic signal variation is essential for understanding genome function. Evo2HiC enables cross–cell type prediction by jointly leveraging both the genome sequence and Hi-C contact maps. In particular, we calculate attribution scores for each base pair in the DNA sequence conditioned on the Hi-C contact maps, and identify motifs with enriched attribution scores for each cell type (see **Methods** for details).

We applied our cell type–specific interpretability framework to two human cell lines: H1ESC and GM12878 (**Fig. 5a**). By using independent gene expression data from the ENCODE project [29], we computed a differential expression score (Δ log(expression)) of each motif between two cell lines. We observed that cell type–specific motifs exhibited substantially higher differential expression compared to shared motifs (**Fig. 5b,c**). As an additional validation, our framework identified the CTCF motif as a shared motif across both cell types, consistent with its well-established role in chromatin organization (**Fig. 5d–f**).

For cell type–specific motifs, we further observed higher expression levels in their corresponding cell lines. To further quantitatively assess motif relevance, we introduced an attribution enrichment score (**Fig. 5c**), which measures the proportion of motif-associated regions receiving high attribution scores relative to background regions. All cell type–specific motifs exhibited strong attribution enrichment, demonstrating Evo2HiC 's ability to precisely identify regulatory motifs that are both structurally and transcriptionally relevant in a cell type–specific manner. As a control, the K562 cell line showed neither high expression nor elevated attribution scores for motifs specific to H1ESC and GM12878, confirming the specificity of our findings.

Moreover, the discovered cell type–specific motifs align well with previous studies. For instance, IRF family motifs were specifically enriched in GM12878, consistent with their established role in mediating cell type–specific transcription factor binding [30]. RARG binding motifs were identified in H1ESC, consistent with prior work linking RARG to retinoic acid–induced enhancer activation and increased H3K27ac 8n embryonic stem cells [31]. These observations support the biological relevance of the motifs recovered by our framework. Finally, Evo2HiC can provide more fine-grained interpretability by localizing active motif binding sites (**Fig. 5g**). By contrasting epigenomic signal predictions between contact matrices with and without the chromatin loop, we can distinguish active from inactive binding sites, further elucidating the context-dependent activity of regulatory motifs.

Evo2HiC demonstrates strong generalizability across 177 species

As Evo 2 is pretrained on multiple species, we hypothesized that the genomic sequence encoder distilled from Evo 2 using only human sequences might generalize to other species. To verify this, we evaluated Evo2HiC on the Hi-C resolution enhancement task, which takes genomic sequences and low-coverage Hi-C matrices as input and aims to reconstruct the corresponding high-resolution contact matrices. In this setting, a U-Net–based decoder was used to generate Hi-C contact patches from the combined embeddings of genomic sequences and Hi-C patches.

We first evaluated performance on 14 human contact maps and observed that Evo2HiC achieved the best results across all four evaluation metrics (**Fig. 6a**), demonstrating the effectiveness of jointly leveraging sequence and contact information for Hi-C resolution enhancement. To assess cross-species generalization, we next applied Evo2HiC to 11 mouse contact maps, where the mouse genomic sequence embeddings were generated without access to Evo 2 mouse embeddings (**Fig. 6b, Supplementary Figure 8**). Nevertheless, Evo2HiC still achieved the best performance, with 5.9% improvement in Spearman correlation and 2.5% improvement in SSIM over the best competing method, confirming its ability to generalize beyond the species used for distillation.

Finally, we extended the evaluation to the DNA Zoo dataset by collecting genomic sequences and Hi-C contact matrices from 177 species (**Fig. 6c, Supplementary Figure 9**). Without utilizing Evo 2 embeddings for these species, Evo2HiC still outperformed the state-of-the-art HiCARN2 model on 157 out of 177 species, further supporting its robustness and generalizability across diverse species.

bioRxiv preprint doi: https://doi.org/10.1101/2025.11.18.689171; this version posted November 19, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

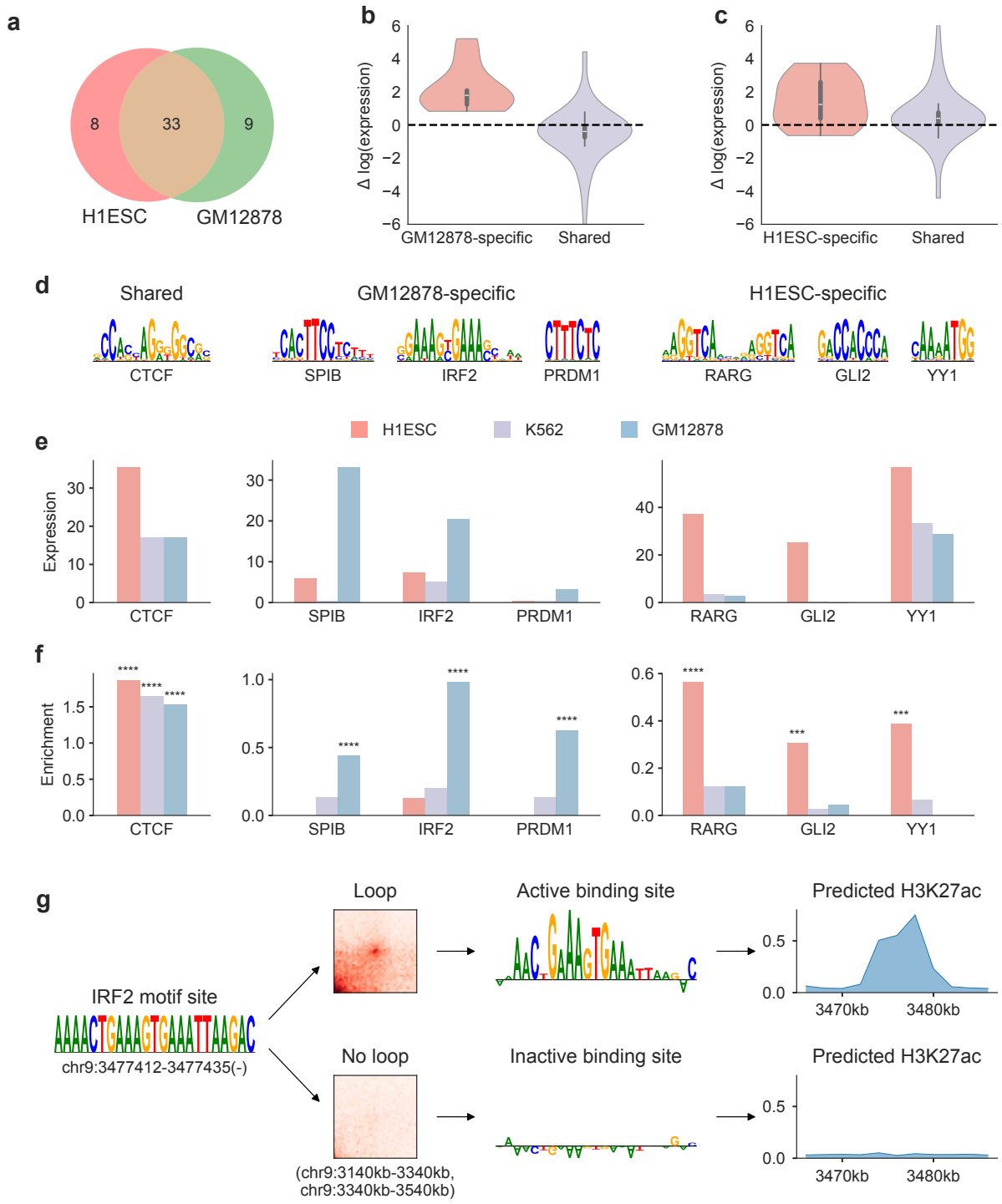


Figure 5: **Identifying cell type-specific motifs for interpreting epigenomic signal changes. a,** Number of shared and cell type-specific motifs identified for GM12878 and K562 by Evo2HiC. **b,c,** Log expression change between cell type-specific motifs and shared motifs on GM12878 (**b**) and H1ESC (**c**). **d,** Shared and cell type-specific motifs identified for GM12878 and H1ESC by Evo2HiC. **e,f,** Gene expression (**e**) and attribution enrichment (**f**) of different motifs across two cell types. Attribution enrichment is calculated by the SHAP attribution score of that motif against random sequences. CTCF is a shared motif. SPIB, IRF2, PRDM1 are identified GM12878-specific motifs. RARG, GLI2, YY1 are identified H1ESC-specific motifs. **g,** Evo2HiC reveals the active binding site for a cell type-specific motif for more fine-grained interpretation of epigenomic signal changes.

bioRxiv preprint doi: https://doi.org/10.1101/2025.11.18.689171; this version posted November 19, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

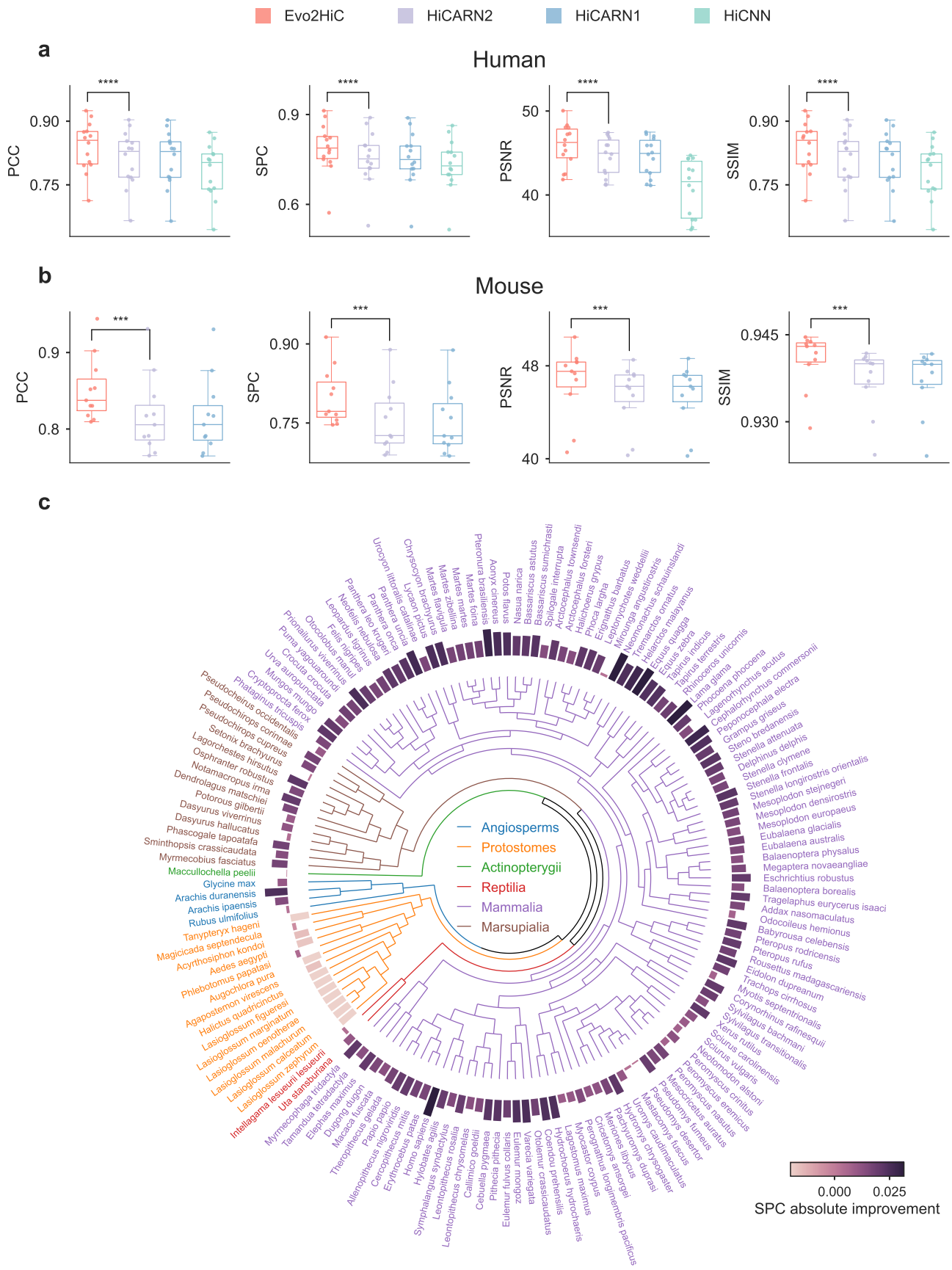


Figure 6: **Evo2HiC enables cross-species analysis on the DNA Zoo data. a,b,** Comparison on resolution enhancement in human (**a**) and mouse (**b**) across four metrics. **c,** Comparison on resolution enhancement on 177 species using DNA Zoo. Bar heights indicate the relative improvement over HiCARN2 by Evo2HiC in terms of Spearman correlation.

bioRxiv preprint doi: https://doi.org/10.1101/2025.11.18.689171; this version posted November 19, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Discussion

We have presented Evo2HiC, a lightweight and effective computational framework for jointly analyzing Hi-C contact matrices and genomic sequences. The key idea of Evo2HiC is to distill a large-scale DNA foundation model (Evo 2) into a much smaller encoder, while guiding the distillation with Hi-C data to retain genomic features critical for three-dimensional genome analysis. We systematically evaluated Evo2HiC across three major applications: (1) predicting Hi-C contact matrix from genomic sequences, (2) predicting epigenomic profiles from joint Hi-C and sequence embeddings, and (3) enhancing Hi-C resolution by integrating Hi-C and sequence. The superior performance across all tasks confirms the effectiveness of our joint modeling framework. Moreover, Evo2HiC enables interpretable analysis by identifying cell type–specific sequence motifs that explain the observed epigenomic profiles, offering mechanistic insights into how 3D genome architecture regulates essential genomic functions.

Despite the promising results of Evo2HiC, several limitations remain that we aim to address in future work. First, beyond Evo 2, a growing number of DNA foundation models [23, 24, 25] have demonstrated strong performance across diverse genomic tasks. We plan to investigate how distilling a compact encoder from multiple foundation models could further enhance the robustness and generalizability of our model. Second, while our results on predicting epigenomic signals has demonstrated potential for studying gene expression, we intend to extend it to directly predict RNA-seq profiles [32], which could provide deeper insights into how 3D chromatin organization influences gene expression. Third, in the current work we extracted embeddings from Evo 2 by freezing its parameters, as the model's size makes fine-tuning computationally challenging. We plan to develop efficient fine-tuning strategies [33, 34] that enable adapting Evo 2 with Hi-C data, potentially unlocking its full capacity for chromatin structure analysis.

bioRxiv preprint doi: https://doi.org/10.1101/2025.11.18.689171; this version posted November 19, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Methods

Overview of Evo2HiC

Evo2HiC has three main components: the Hi-C encoder, the DNA encoder, and task-specific heads for downstream applications.

The Hi-C encoder employs a CNN-based architecture, comprising an initial cross-embedding layer and two residual blocks. The cross-embedding layer applies three convolutional operations to the Hi-C matrices with kernel sizes of 3, 7, and 15, producing feature maps of dimensions 64, 32, and 32, respectively. These are concatenated to form a 128-dimensional Hi-C embedding. Subsequently, two residual blocks are incorporated to enhance the encoder's representational capacity. In addition to standard residual blocks, ours are modified to adapt to multiple resolutions of Hi-C, allowing for the capture of chromatin architecture at different scales: the input embedding is pooled at multiple scales (with pooling sizes of 1, 2, 4, and 5) before applying 3×3 convolutional layers. The resulting multi-resolution embeddings are then concatenated and projected back to a 128-dimensional representation.

The DNA encoder adopts a two-stage architecture. In the first stage, a 1D convolutional encoder processes the one-hot encoded DNA sequence together with the DNA mappability track to generate 1D embeddings. The architecture of this stage is adapted from Orca [28] and consists of seven stacked 1D residual blocks with kernel sizes of 9 and channel dimensions of 64, 96, 128, 128, 128, 128, and 128, respectively. Each residual block is preceded by a max-pooling layer with pooling sizes of 1, 4, 4, 5, 5, 5, and 1, progressively downsampling the input sequence to produce embeddings at an effective resolution of 2 kb. In the second stage of the DNA encoder, the 1D embeddings corresponding to the row and column indices of the Hi-C matrix are combined by summation and passed through a multi-resolution module identical to that used in the Hi-C encoder. The resulting representations are then projected back into a 128-dimensional embedding space.

Evo 2 embeddings

To reduce computational cost, we precomputed Evo 2 embeddings on reference genomes (hg38 for training and evaluation, and mm10 for evaluation only). Specifically, we performed inference using the official Evo 2-7B checkpoint, extracted embeddings from layer 27, and averaged them within non-overlapping 2 kb bins, following the protocol described in the Evo 2 paper. To further improve efficiency, inference was conducted on 60 kb subsequences with a 50 kb stride along the genome. Because Evo 2 was designed with a causal mask (i.e., each token can only see tokens before it), we discard the initial 10 kb of each subsequence to avoid edge effects. Inference was performed on both DNA strands, and the resulting strand-specific embeddings were concatenated to obtain the final Evo 2 representations of dimension 8192.

Hi-C data

For pretraining, we collected high read-depth Hi-C contact maps from a wide range of sources. The dataset comprises 149 Hi-C maps curated from ENCODE [29] and 4DN [35], including experiments based on *in situ* Hi-C, intact Hi-C, and Micro-C protocols. We performed training, validation, and testing splits across both species, cell lines, and chromosomes. Specifically, the dataset includes 138 human Hi-C maps, of which 124 maps from 46 cell lines were used for training and 14 maps from 6 cell lines were held out for testing. Additionally, 11 mouse Hi-C maps from 4 cell lines were used exclusively for testing. For all contact maps, chromosome 8 was reserved for validation, while chromosomes 9 and 10 were left out for testing.

Pretraining

During pretraining, two complementary distillation processes were jointly applied to inject prior knowledge into the encoders. The first process is sequence distillation, where Evo 2 acts as the sequence teacher and the DNA encoder as the student, guiding the encoder to capture sequence-level patterns such as transcription factor motifs. The second process is structure distillation, in which the Hi-C encoder extracts structural knowledge and serves as the structure teacher, transferring knowledge of 3D chromatin organization to help the DNA encoder learn structurally relevant representations.

bioRxiv preprint doi: https://doi.org/10.1101/2025.11.18.689171; this version posted November 19, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

In sequence distillation, we adopt a contrastive learning framework to inject Evo 2's knowledge into the DNA encoder. Specifically, embeddings derived from the same genomic sequence by Evo 2 and the DNA encoder are treated as positive pairs, while embeddings from different sequences are treated as negative pairs. This setup encourages the DNA encoder to learn an embedding space consistent with that of Evo 2. At each training step on a single GPU, we randomly sample six 200 kb regions from the human genome training chromosomes. The DNA encoder produces 1D DNA embeddings for every non-overlapping 2 kb bin within these regions, with 20 kb trimmed from both edges to avoid boundary artifacts, yielding 480 embeddings of dimension 128. The corresponding pre-computed Evo 2 embeddings for these 2 kb bins are then retrieved. To enrich the pool of negative samples, we additionally sample 20,480 bins across the genome and obtain their Evo 2 embeddings. This procedure results in 10,060,800 sample pairs in total, of which 480 are positive. To perform the contrastive alignment, additional linear projection layers are appended to both the DNA encoder and Evo 2 embeddings, mapping them into a shared 512-dimensional space. The contrastive distillation loss follows the formulation of SigLIP [36] and is defined as

$$\mathcal{L}_{seq} = \frac{1}{N_{seq}} \sum_{i=1}^{N_{seq}} \sum_{j=1}^{M_{seq}} \log \frac{1}{1 + e^{z_{ij}(-t_{seq} x_i y_j + b_{seq})}},$$

where N_{seq} and M_{seq} denote the numbers of DNA encoder and Evo 2 embeddings, respectively; x_i and y_j represent the i-th DNA encoder and j-th Evo 2 embedding; z_{ij} equals 1 for positive pairs and -1 for negative pairs; and t_{seq} and b_{seq} are learnable temperature and bias parameters that modulate the contrastive scale.

In the structure distillation stage, we adopt a similar contrastive learning framework to transfer structural knowledge, such as chromatin loops and topologically associating domains (TADs), from Hi-C contact maps into the DNA encoder. Specifically, for each 2 kb $\times$ 2 kb pixel, the Hi-C encoder extracts an embedding from the Hi-C patch centered at that pixel, while the DNA encoder computes a corresponding 2D embedding using the DNA sequences of the row and column bins defining the patch. Within the contrastive framework, Hi-C and 2D DNA embeddings from the same pixel are treated as positive pairs, whereas embeddings from different pixels are treated as negative pairs. At each training step on a single GPU, we randomly sample six 200 kb $\times$ 200 kb submatrices at 2 kb resolution from the human Hi-C contact map within a 500 kb diagonal window, excluding the validation and test chromosomes. For each submatrix, eight Hi-C contact maps are randomly selected, and the Hi-C encoder computes embeddings for every pixel across these maps. The embeddings from these eight contact maps are then averaged to obtain the final Hi-C embeddings for pixels. The DNA encoder generates the corresponding 2D DNA embeddings for all pixels, as described in the previous section. To avoid boundary effects, embeddings at both 20 kb borders are discarded. This procedure yields a total of 1,474,560,000 sample pairs, among which 38,400 are positive. Similarly, the contrastive distillation loss follows the formulation of SigLIP and is defined as

$$\mathcal{L}_{struct} = \frac{1}{N_{struct}} \sum_{i=1}^{N_{struct}} \sum_{j=1}^{N_{struct}} \log \frac{1}{1 + e^{w_{ij}(-t_{struct} u_i v_j + b_{struct})}},$$

where N_{struct} denotes the number of Hi-C and 2D DNA embeddings, u_i and v_j represent the i-th Hi-C and j-th 2D DNA embedding, respectively, w_{ij} equals 1 for positive pairs and -1 for negative pairs, and t_{stru} and b_{stru} are learnable temperature and bias parameters that control the contrastive scaling.

The final pretraining objective is defined as

$$\mathcal{L} = \mathcal{L}_{seq} + \mathcal{L}_{struct}.$$

Using this loss, we trained the model on four NVIDIA A100 GPUs with Adam optimizer, an initial learning rate of 10^{-4}, and a total of 50,000 training steps, which was determined based on convergence on the validation set.

Details on measuring embedding space similarity and retrieval

After distilling knowledge from Evo 2 and Hi-C, we evaluated whether the DNA encoder learned an embedding space consistent with Evo 2. To perform this analysis in the original Evo 2 space, we fitted a linear

bioRxiv preprint doi: https://doi.org/10.1101/2025.11.18.689171; this version posted November 19, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

transformation from the 128-dimensional 1D DNA embeddings to the 8192-dimensional Evo 2 embeddings using a least-squares objective on chromosome 9. The learned transformation was then applied to 1D DNA embeddings from chromosome 10 of the human genome (hg38) and chromosome 4 of the mouse genome (mm10). Finally, we quantified the similarity between embedding spaces by comparing pairwise similarities among DNA bins in the Evo 2 space and in our DNA encoder's transformed space. 5000 pairs of bins were sampled, and the similarity between each pair in the two spaces was computed. The Spearman correlation coefficient (SPC) between these two lists of distances was reported as the final measure of embedding-space similarity.

Next, we assessed whether the encoders successfully captured structural knowledge by evaluating their ability to retrieve Hi-C pixels directly from the corresponding DNA sequences. For computational efficiency, retrieval was restricted to Hi-C pixels located at the same distance from the diagonal within the test chromosomes. We computed the cosine similarity between Hi-C embeddings and 2D DNA embeddings. We reported the top-1, top-5, and top-10 recall (where top-k recall indicates that the positive pair appears within the k most similar pairs), the average rank of the positive pairs, and the similarity change between positive pairs and negative pairs.

Predicting Hi-C contact maps

For Hi-C contact map prediction, we used 4 kb–resolution Micro-C data from the H1ESC and HFF cell lines, normalized with SCALE normalization [37]. For both cell lines, chromosome 8 was reserved for validation, and chromosomes 9 and 10 were used for testing. Notably, the H1ESC cell line was entirely held out during pretraining, and the validation and test chromosomes of the HFF cell line were also excluded from the pretraining data.

We designed a U-Net–style architecture [38] as the decoder for this task, which takes 2D DNA embeddings as input and outputs the target Hi-C submatrices. In particular, this U-Net-style architecture is an encoder-decoder architecture with skip connections, where the encoder utilizes three residual blocks to extract features at different granularities, while the decoder employs another three residual blocks to generate the desired output [38]. We fine-tuned Evo2HiC separately for the two cell lines. During fine-tuning, we randomly sampled submatrices of size 1120 kb $\times$ 1120 kb along the diagonal of the Hi-C matrix. Each GPU processed a batch of seven submatrices, resulting in a total batch size of 28 across four NVIDIA A100 GPUs. The model was trained using the mean squared error (MSE) loss:

$$\mathcal{L}_{MSE} = \frac{1}{N}\|p - t\|_2^2,$$

where p and t denote the predicted and target Hi-C submatrices, respectively, and N is the submatrix size. Data augmentation was applied by randomly flipping both the Hi-C submatrices and the corresponding DNA sequences with a probability of 0.5, where flipping refers to reversing the DNA sequences to their reverse-complement strand and mirroring the Hi-C submatrix along both row and column axes. In addition, the DNA sequences were randomly shifted by 100 bp. The model was trained using the Adam optimizer with an initial learning rate of 10^{-4} for a total of 20,000 steps. For evaluation, we applied the model to predict 1 Mb $\times$ 1 Mb sliding windows along the diagonal of the Hi-C matrices with a stride of 0.5 Mb. Because the model was trained to predict 1120 kb $\times$ 1120 kb submatrices, we trimmed 80 kb from the top and left borders and 40 kb from the bottom and right borders to match the target window size.

We next prepared Orca and Evo 2 as baselines for Hi-C contact map prediction. For Orca, since our evaluation uses the same data sources, we directly downloaded the Orca predictions from the official repository and followed the provided instructions to convert them back into Hi-C contact maps. For Evo 2, we replaced the 2D DNA embeddings with the pre-computed Evo 2 embeddings and used the same decoder architecture and fine-tuning objective to predict Hi-C maps from Evo 2 embeddings.

To evaluate model performance, we followed a widely used scheme based on distance-stratified correlation and insulation score correlation [39]. For distance-stratified correlation, pixels in the Hi-C matrices were grouped by their distance to the diagonal. The Spearman correlation coefficient (SPC), Pearson correlation coefficient (PCC), and Peak signal-to-noise ratio (PSNR) were computed between predicted and target values within each group. For insulation score correlation, we used cooltools [40] to calculate the insulation score for each bin in the submatrices and reported the Pearson correlation coefficient (PCC) and Peak signal-to-noise ratio (PSNR) between the predicted and target insulation profiles.

bioRxiv preprint doi: https://doi.org/10.1101/2025.11.18.689171; this version posted November 19, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Predicting epigenomic profiles

For epigenomic signal prediction, we collected DNase-seq, ChIP-seq CTCF, ChIP-seq H3K27ac, ChIP-seq H3K27me3, and ChIP-seq H3K4me3 data for the GM12878, H1ESC, and K562 cell lines from ENCODE (accession codes in **Supplementary Table 1**). As in pretraining, chromosome 8 was reserved for validation, and chromosomes 9 and 10 were used for testing. The entire K562 cell line was additionally held out for testing to assess the model's ability to generalize to unseen cell types. All epigenomic signals were binned at 2 kb resolution. To normalize signal scales across cell lines, we computed the 90th percentile value (q) of the signal distribution and used $20 \times q$ as a cutoff threshold, normalizing signal intensities to the range [0,1]. We confirmed that in every cell line, fewer than 1% of bins were truncated by this procedure, indicating that nearly all information was preserved.

We next designed a transformer-based decoder to predict these epigenomic signals from multimodal embeddings. The decoder takes 1D DNA embeddings and Hi-C embeddings as input to predict epigenomic signals. Specifically, the Hi-C embeddings were computed with the Hi-C encoder for a 320 kb $\times$ 2 Mb region centered along the diagonal and then averaged over the 2 Mb dimension to obtain 1D representations. These 1D Hi-C embeddings were concatenated with the 1D DNA embeddings to form the input to the transformer decoder. The decoder consists of 11 transformer encoder layers with a base hidden dimension of 256 and employs Rotary Position Embedding (RoPE) [41] to encode positional information.

For the finetuning process, the model predicts epigenomic signals within a 320 kb window, using as input a 320 kb DNA sequence together with a 320 kb $\times$ 2 Mb Hi-C submatrix centered along the diagonal. The fine-tuning loss combines cosine similarity and mean squared error (MSE), defined as

$$\mathcal{L}_{MSE} = \frac{1}{N}\|p - t\|_2^2,$$
$$\mathcal{L}_{COS} = \frac{p \cdot t}{\|p\|_2 \|t\|_2},$$
$$\mathcal{L} = \mathcal{L}_{MSE} + \mathcal{L}_{COS},$$

where p and t denote the prediction and target, respectively, and N is the size of target sample. The model was fine-tuned with a total batch size of 16 across four GPUs using the Adam optimizer, an initial learning rate of 10^{-4}, and a total of 50,000 steps.

We evaluated Evo2HiC against two baseline models. The first baseline, a Hi-C–only model, uses the same architecture as Evo2HiC but excludes DNA sequences from the input. The second baseline, an Evo 2–only model, employs the same decoder but replaces DNA sequences with pre-computed Evo 2 embeddings and removes Hi-C input. Because the Evo 2–only model lacks cell line–specific structural information, we trained separate models for each cell line, including the leave-out cell line. Model performance was evaluated using Pearson and Spearman correlation coefficients (PCC and SPC) between the predicted and target epigenomic signals over the test chromosomes.

Identifying cell type-specific motifs

We identified cell type–specific motifs by attributing epigenomic signal variation to individual base pairs. Specifically, we used the DeepLiftShap [42] algorithm implemented in tangermeme [43] to compute attribution scores for each base pair in the DNA sequence, conditioned on the Hi-C contact map of a given cell type. DeepLiftShap was applied to every 320 kb window along the test chromosomes with a stride of 240 kb, excluding windows containing unknown bases. We next obtained a precomputed, genome-wide catalog of motif matches [44]. For each motif, we assessed whether its matching sites were associated with higher attribution scores compared to the local background. A high-attribution site was defined as one whose average attribution score exceeded the chromosome-wide mean by at least five standard deviations. The local background was defined as the flanking region extending ten times the motif length on each side. We then quantified the fold enrichment in the proportion of motif sites with high attribution scores relative to their local background and computed the corresponding p-values using Welch's z-test. To control for multiple testing, we applied the Benjamini–Hochberg false discovery rate (BH-FDR) procedure to all motifs within each cell type. Motifs enriched in a cell type were defined as those with at least 80 high-attribution sites and a BH-FDR–adjusted p-value below 10^{-3}. Cell-type–specific motifs were further defined as those

bioRxiv preprint doi: https://doi.org/10.1101/2025.11.18.689171; this version posted November 19, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

enriched in one cell type but not in any other. We additionally identified cell type–specific motif clusters based on a previously reported motif-similarity analysis [44]. Specifically, cell type-specific motif clusters are clusters that have at least one enriched motif in a cell but no enriched motif in another. Finally, the cell-type–specific motifs and motif clusters were cross-validated using RNA-seq data from ENCODE (accession codes in **Supplementary Table 2**), where transcript-per-million (TPM) values served as an external indicator of motif activity across cell types. For motif clusters, the TPMs of all enriched motifs in the cluster are summed as the TPM for the cluster.

Resolution enhancement across different species

For resolution enhancement, Evo2HiC takes DNA sequences and low-coverage Hi-C matrices as input to reconstruct high-coverage Hi-C contact maps. The model adopts a U-net–style decoder, similar to the one used for predicting Hi-C contact maps. We fine-tuned Evo2HiC on our pretraining dataset using the same data split, simulating low-coverage inputs through binomial downsampling with a ratio of 1/16. The training set includes only human cell lines. To better leverage DNA sequence information, we trained the model on Hi-C matrices at 2 kb resolution. In addition, we introduced a multi-resolution mean squared error (MSE) loss to mitigate sparsity at this fine resolution, defined as

$$\mathcal{L}_{mr} = \frac{1}{|R|} \sum_{r \in R} \|\text{meanpool}_r(p) - \text{meanpool}_r(t)\|_2^2,$$

where p and t denote the predicted and target Hi-C contact maps, respectively; R denotes the set of resolutions, and in practice $R = \{2\text{ kb}, 4\text{ kb}, 8\text{ kb}, 10\text{ kb}\}$; $\text{meanpool}_r(\cdot)$ denotes average pooling to resolution r. We fine-tuned the model with a total batch size of 128 using the Adam optimizer, an initial learning rate of 10^{-4}, and 200,000 optimization steps.

We first evaluated model performance on the held-out test cell lines from the pretraining dataset, which include six human and four mouse cell lines. Four metrics were used for evaluation: Peak signal-to-noise ratio (PSNR), Structural similarity index measure (SSIM), and distance-stratified Pearson and Spearman correlation coefficients (PCC and SPC). The distance-stratified PCC and SPC are computed within distance-specific groups of bins and averaged across groups. Following prior work [45], all metrics were computed at 10 kb resolution within a 2 Mb genomic distance. To further assess Evo2HiC's cross-species generalization, we obtained Hi-C contact maps and DNA sequences from the DNA Zoo dataset [46]. DNA sequences and Hi-C matrices were paired following the DNA Zoo protocol, resulting in 177 sequence–contact map pairs. Using the downsampled low-coverage Hi-C maps and corresponding DNA sequences, we reconstructed the original high-resolution maps and reported distance-stratified SPC and PCC as the evaluation metrics.

As baselines, we included HiCNN [47], HiCARN1, and HiCARN2 [45], following their original training procedures but replacing their datasets with our pretraining data.

Code Availability

Evo2HiC is accessible at https://github.com/CHNFTQ/Evo2HiC, including the model weights and relevant source code, released under the Apache License 2.0. We include detailed methods and implementation steps in the Methods and Supplementary Information to allow for independent replication.

Acknowledgement

Figure 1 was created using icons adapted from "Various Illustration Vectors" by soco-st, licensed under CC BY 4.0.

bioRxiv preprint doi: https://doi.org/10.1101/2025.11.18.689171; this version posted November 19, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Author Contributions Statement

	TF	XW	ZX	SH	GM	JY	HX	AJ	WSN	SW
Conceptualization	•	•							•	•
Data curation	•	•								
Funding acquisition									•	•
Investigation	•									
Methodology	•	•	•	•	•	•	•	•	•	•
Project administration	•	•							•	•
Software	•	•								
Resources	•								•	•
Supervision									•	•
Visualization	•	•	•				•			
Writing—original draft	•									•
Writing—review & editing	•	•	•	•	•	•	•	•	•	•

References

[1] Eric H Davidson and Douglas H Erwin. "Gene regulatory networks and the evolution of animal body plans". In: *Science* 311.5762 (2006), pp. 796–800.

[2] Michael Levine and Eric H Davidson. "Gene regulatory networks for development". In: *Proceedings of the National Academy of Sciences* 102.14 (2005), pp. 4936–4942.

[3] Isabelle S Peter and Eric H Davidson. "Evolution of gene regulatory networks controlling body plan development". In: *Cell* 144.6 (2011), pp. 970–985.

[4] Kirsten I Bos et al. "A draft genome of Yersinia pestis from victims of the Black Death". In: *Nature* 478.7370 (2011), pp. 506–510.

[5] Bernard Dujon et al. "Genome evolution in yeasts". In: *Nature* 430.6995 (2004), pp. 35–44.

[6] Igor Ulitsky and David P Bartel. "lincRNAs: genomics, evolution, and mechanisms". In: *Cell* 154.1 (2013), pp. 26–46.

[7] Douglas Hanahan and Robert A Weinberg. "Hallmarks of cancer: the next generation". In: *cell* 144.5 (2011), pp. 646–674.

[8] Ludmil B Alexandrov et al. "The repertoire of mutational signatures in human cancer". In: *Nature* 578.7793 (2020), pp. 94–101.

[9] Margaret A Hamburg and Francis S Collins. "The path to personalized medicine". In: *New England Journal of Medicine* 363.4 (2010), pp. 301–304.

[10] Anna Portela and Manel Esteller. "Epigenetic modifications and human disease". In: *Nature biotechnology* 28.10 (2010), pp. 1057–1068.

[11] Teri A Manolio et al. "Finding the missing heritability of complex diseases". In: *Nature* 461.7265 (2009), pp. 747–753.

[12] Amit V Khera et al. "Genetic risk, adherence to a healthy lifestyle, and coronary disease". In: *New England Journal of Medicine* 375.24 (2016), pp. 2349–2358.

[13] 4D Nucleome Consortium et al. "An integrated view of the structure and function of the human 4D nucleome". In: *bioRxiv* (2024), pp. 2024–09.

[14] S. S. P. Rao et al. "A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping". In: *Cell* 59.7 (2014), pp. 1665–1680.

[15] R. A. Beagrie et al. "Complex multi-enhancer contacts captured by genome architecture mapping". In: *Nature* 543.7646 (2017), pp. 519–524.

[16] Sofia A Quinodoz et al. "SPRITE: a genome-wide method for mapping higher-order 3D interactions in the nucleus using combinatorial split-and-pool barcoding". In: *Nature protocols* 17.1 (2022), pp. 36–75.

[17] Maxwell R Mumbach et al. "HiChIP: efficient and sensitive analysis of protein-directed genome architecture". In: *Nature methods* 13.11 (2016), pp. 919–922.

[18] M Jordan Rowley and Victor G Corces. "Organizational principles of 3D genome architecture". In: *Nature Reviews Genetics* 19.12 (2018), pp. 789–800.

bioRxiv preprint doi: https://doi.org/10.1101/2025.11.18.689171; this version posted November 19, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[19] Chin-Tong Ong and Victor G Corces. "CTCF: an architectural protein bridging genome topology and function". In: *Nature Reviews Genetics* 15.4 (2014), pp. 234–246.

[20] Jian Zhou. "Sequence-based modeling of three-dimensional genome architecture from kilobase to chromosome scale". In: *Nature genetics* 54.5 (2022), pp. 725–734.

[21] Geoff Fudenberg, David R Kelley, and Katherine S Pollard. "Predicting 3D genome folding from DNA sequence with Akita". In: *Nature methods* 17.11 (2020), pp. 1111–1117.

[22] R. Schwessinger et al. "DeepC: predicting 3D genome folding using megabase-scale transfer learning". In: *Nature methods* 17.11 (2020), pp. 1118–1124.

[23] Yanrong Ji et al. "DNABERT: pre-trained Bidirectional Encoder Representations from Transformers model for DNA-language in genome". In: *Bioinformatics* 37.15 (2021), pp. 2112–2120.

[24] Zhihan Zhou et al. "Dnabert-2: Efficient foundation model and benchmark for multi-species genome". In: *arXiv preprint arXiv:2306.15006* (2023).

[25] Hugo Dalla-Torre et al. "Nucleotide Transformer: building and evaluating robust foundation models for human genomics". In: *Nature Methods* 22.2 (2025), pp. 287–297.

[26] Eric Nguyen et al. "Sequence modeling and design from molecular to genome scale with Evo". In: *Science* 386.6723 (2024), eado9336.

[27] Garyk Brixi et al. "Genome modeling and design across all domains of life with Evo 2". In: *BioRxiv* (2025), pp. 2025–02.

[28] J. Zhou. "Sequence-based modeling of three-dimensional genome architecture from kilobase to chromosome scale". In: *Nature Genetics* 54 (2022), pp. 725–734.

[29] ENCODE Project Consortium. "The ENCODE (ENCyclopedia Of DNA Elements) Project". In: *Science* 306.5696 (2004), pp. 636–640.

[30] Akshay Kakumanu et al. "Deconvolving sequence features that discriminate between overlapping regulatory annotations". In: *PLoS computational biology* 13.10 (2017), e1005795.

[31] Vasundhra Kashyap et al. "RARγ is essential for retinoic acid induced chromatin remodeling and transcriptional activation in embryonic stem cells". In: *Journal of cell science* 126.4 (2013), pp. 999–1008.

[32] Ali Mortazavi et al. "Mapping and quantifying mammalian transcriptomes by RNA-Seq". In: *Nature methods* 5.7 (2008), pp. 621–628.

[33] Neil Houlsby et al. "Parameter-efficient transfer learning for NLP". In: *International conference on machine learning*. PMLR. 2019, pp. 2790–2799.

[34] Edward J Hu et al. "Lora: Low-rank adaptation of large language models." In: *ICLR* 1.2 (2022), p. 3.

[35] J. Dekker et al. "The 4D Nucleome Project". In: *Nature* 549 (2017), pp. 219–226.

[36] Xiaohua Zhai et al. "Sigmoid Loss for Language Image Pre-Training". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. Oct. 2023, pp. 11975–11986.

[37] N. C. Durand et al. "Juicer Provides a one-click system for analyzing loop-resolution Hi-C experiments". In: *Cell Systems* 3.1 (2016), pp. 95–98.

[38] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation". In: *International Conference on Medical image computing and computer-assisted intervention*. Springer. 2015, pp. 234–241.

[39] Jimin Tan et al. "Cell-type-specific prediction of 3D chromatin organization enables high-throughput in silico genetic screening". In: *Nature Biotechnology* 41.8 (2023), pp. 1140–1150.

[40] Open2C et al. "Cooltools: enabling high-resolution Hi-C analysis in Python". In: *PLOS Computational Biology* 20.5 (2024), e1012067.

[41] Jianlin Su et al. "Roformer: Enhanced transformer with rotary position embedding". In: *Neurocomputing* 568 (2024), p. 127063.

bioRxiv preprint doi: https://doi.org/10.1101/2025.11.18.689171; this version posted November 19, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[42] Scott M Lundberg and Su-In Lee. "A unified approach to interpreting model predictions". In: *Advances in neural information processing systems* 30 (2017).

[43] Jacob Schreiber. "tangermeme: A toolkit for understanding cis-regulatory logic using deep learning models". In: *bioRxiv* (2025), pp. 2025–08.

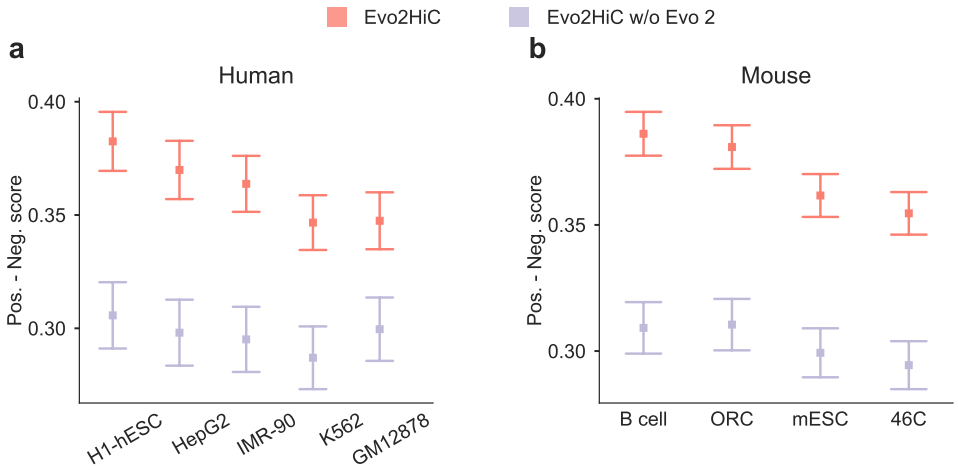
[44] Jeff Vierstra et al. "Global reference mapping of human transcription factor footprints". In: *Nature* 583.7818 (2020), pp. 729–736.

[45] Parker Hicks and Oluwatosin Oluwadare. "HiCARN: resolution enhancement of Hi-C data using cascading residual networks". In: *Bioinformatics* 38.9 (2022), pp. 2414–2421.

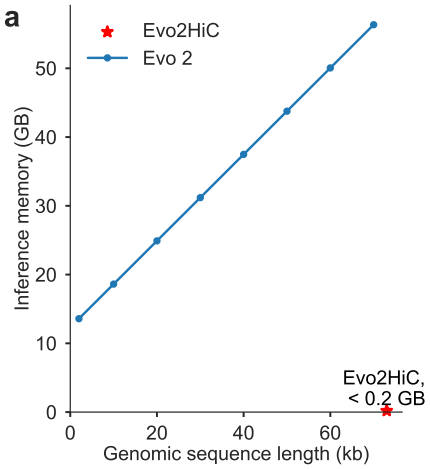
[46] Claire Hoencamp et al. "3D genomics across the tree of life reveals condensin II as a determinant of architecture type". In: *Science* 372.6545 (2021), pp. 984–989.

[47] Tong Liu and Zheng Wang. "HiCNN: a very deep convolutional neural network to better enhance the resolution of Hi-C data". In: *Bioinformatics* 35.21 (2019), pp. 4222–4228.

bioRxiv preprint doi: https://doi.org/10.1101/2025.11.18.689171; this version posted November 19, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

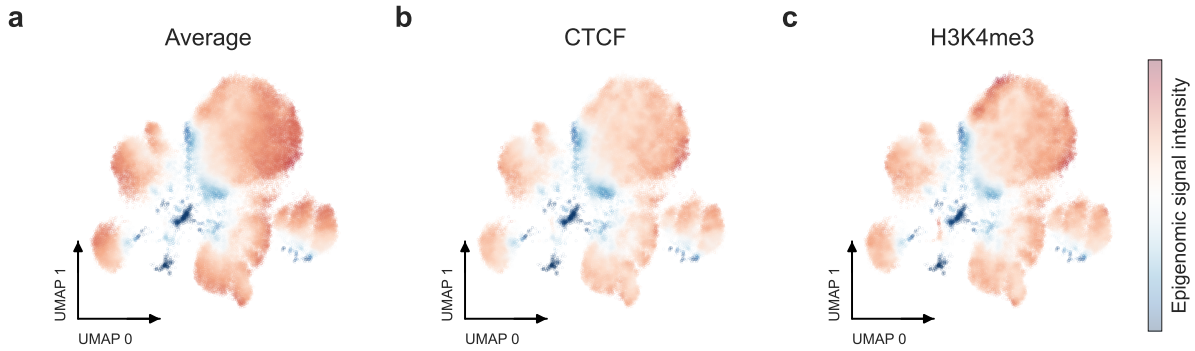


Supplementary Figure 1: **Evaluation of the embedding space of Evo2HiC. a,b** Comparison between Evo2HiC and the Evo2HiC without Evo 2 on retrieving the most similar Hi-C patch using the corresponding genomic sequence on human **(a)** and mouse **(b)**. y-axis is the gap between positive pairs and negative pairs.



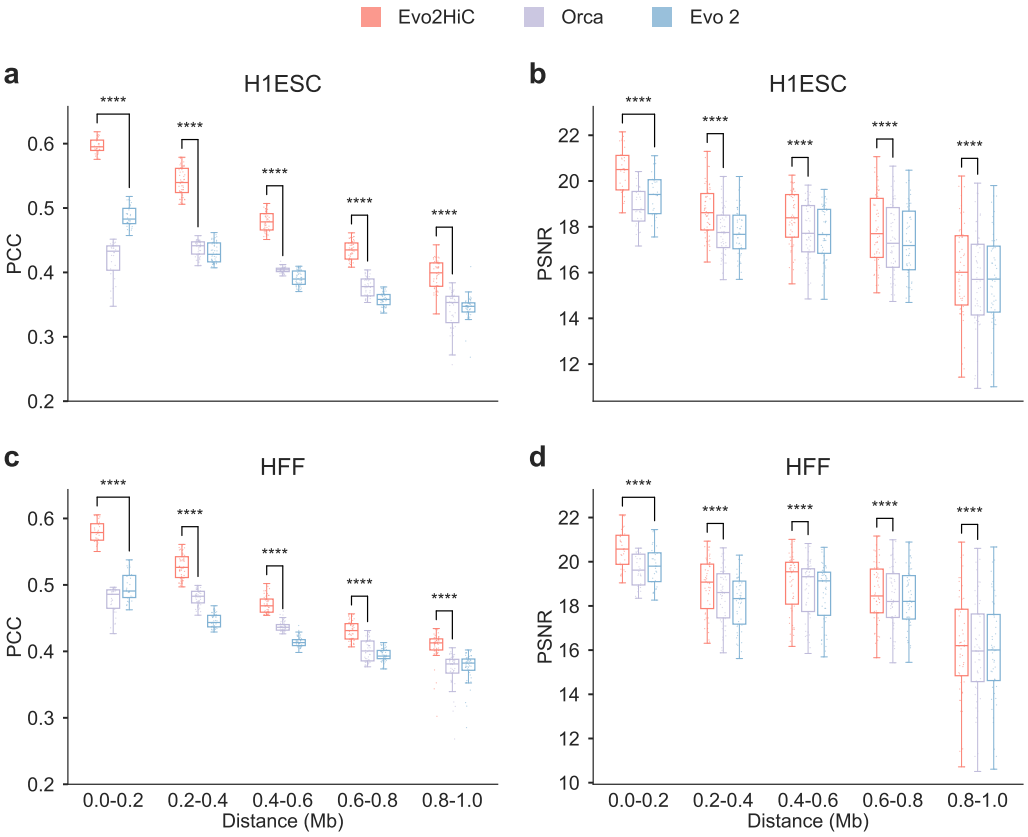
Supplementary Figure 2: **Comparison of the inference memory.** Memory cost for the inference of one genomic sequence with different lengths between Evo 2 and Evo2HiC.

bioRxiv preprint doi: https://doi.org/10.1101/2025.11.18.689171; this version posted November 19, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

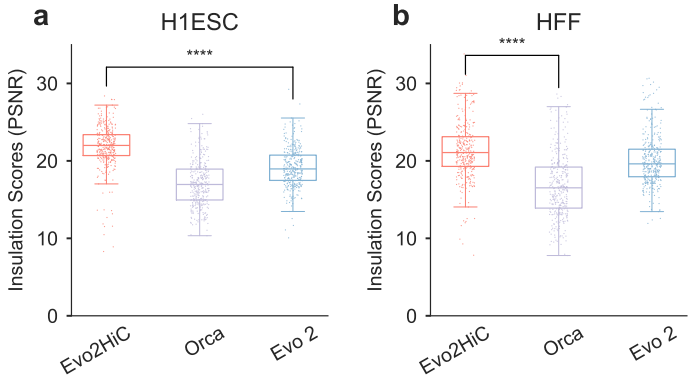


Supplementary Figure 3: **UMAP visualization.** UMAP visualization of the embeddings of genomic sequences by Evo2HiC. **a** is the average epigenomic signal over CTCF, DNase, H3K27ac, H3K27me3, and H3K4me3. **b** is the epigenomic signal of CTCF and **c** is H3K4me3.

bioRxiv preprint doi: https://doi.org/10.1101/2025.11.18.689171; this version posted November 19, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

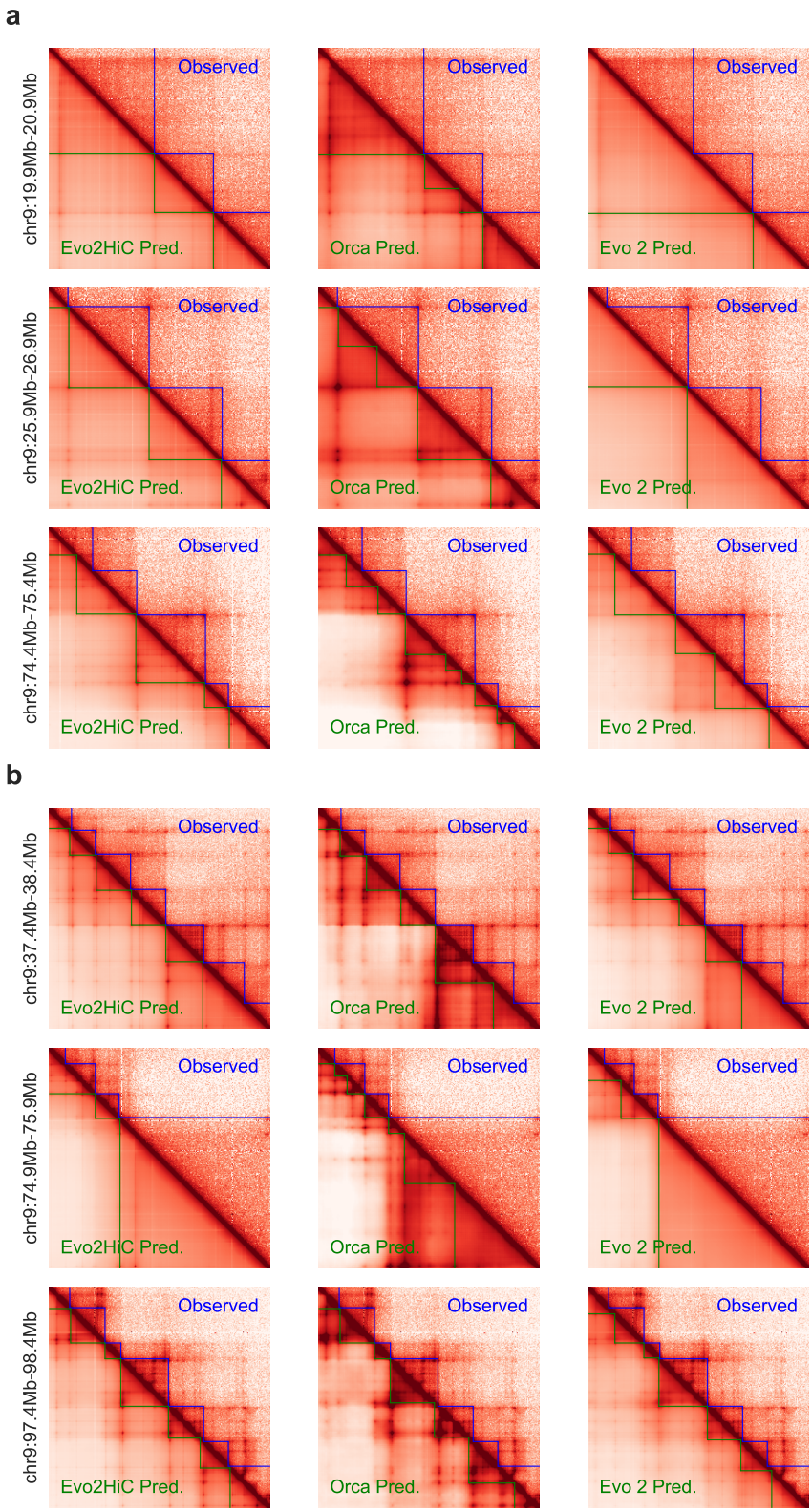


Supplementary Figure 4: **Distance-stratified metrics on genome structure prediction from sequence. a–d,** Comparison among Evo2HiC, Orca, and Evo 2 for predicting Hi-C contact maps from genomic sequence input. Metrics are Pearson correlation (PCC) **(a, c)** and Peak signal-to-noise ratio (PSNR) on Hi-C contact counts **(b, d)**. **a,b** are results on the H1ESC cell line and **c,d** are results on the HFF cell line.



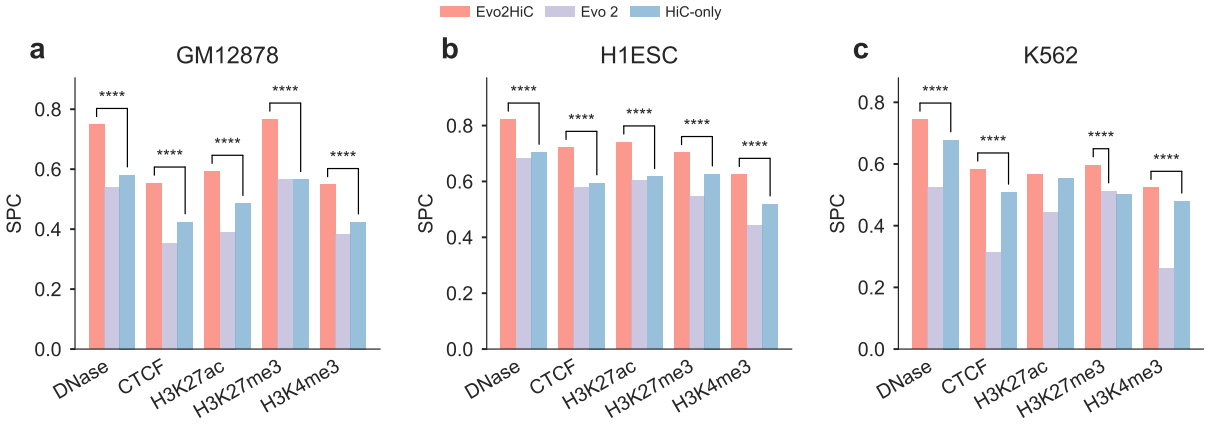
Supplementary Figure 5: **Genome structure prediction from sequence evaluated using insulation score. a–b,** Comparison among Evo2HiC, Orca, and Evo 2 for predicting Hi-C contact maps from genomic sequence input in terms of the Peak signal-to-noise ratio (PSNR) of insulation score on the H1ESC cell line (**a**) and the HFF cell line (**b**).

bioRxiv preprint doi: https://doi.org/10.1101/2025.11.18.689171; this version posted November 19, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

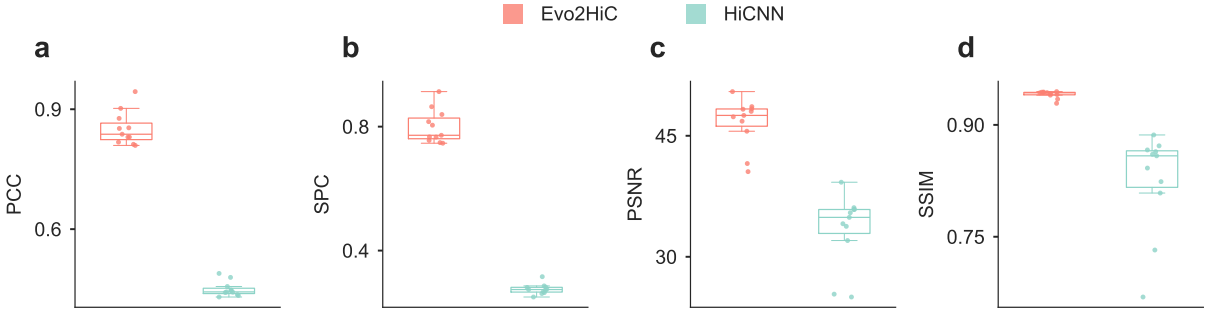


Supplementary Figure 6: **Visualization of genome structure prediction from sequence. a,b,** Visualization of predicted and observed Hi-C contact maps for the H1ESC cell line (**a**) and the HFF cell line (**b**). In each visualization, the lower-left corner shows the predicted matrices from Evo2HiC, Orca, and Evo 2, respectively, while the top-right corner shows the observed matrix.

bioRxiv preprint doi: https://doi.org/10.1101/2025.11.18.689171; this version posted November 19, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

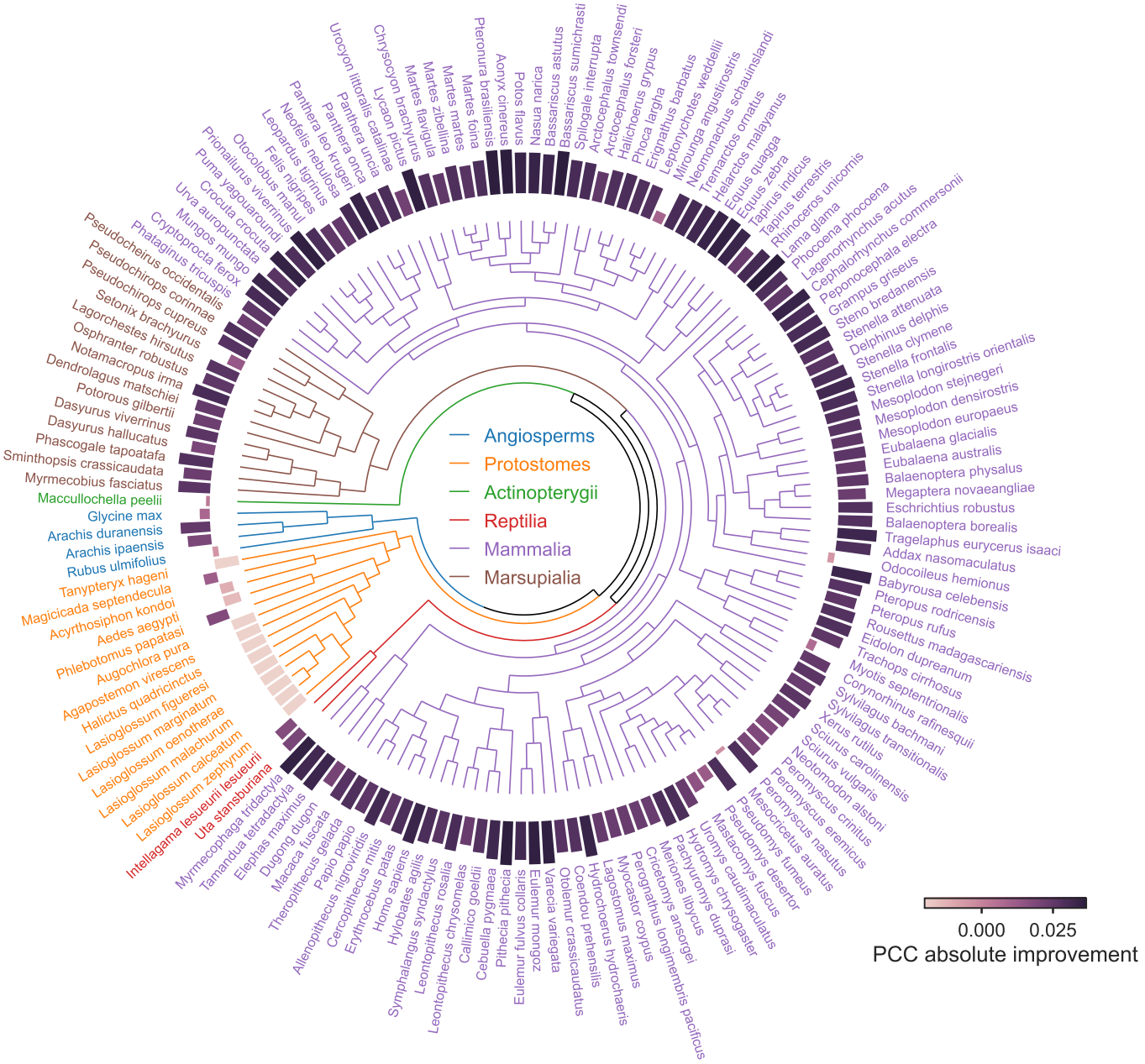


Supplementary Figure 7: **Evaluation on predicting epigenomic signals. a,b,** Comparison on predicting epigenomic signals in the cross chromosome setting on GM12878 (**a**) and H1ESC (**b**). **c,** Comparison on predicting epigenomic signals in the cross cell line and cross chromosome setting on K562.



Supplementary Figure 8: Comparison on resolution enhancement in mouse (**b**) across four metrics between Evo2HiC and HiCNN.

bioRxiv preprint doi: https://doi.org/10.1101/2025.11.18.689171; this version posted November 19, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



Supplementary Figure 9: **Comparison on resolution enhancement on 177 species using DNA Zoo.** Bar heights indicate the relative improvement over HiCARN2 by Evo2HiC in terms of Pearson correlation.

bioRxiv preprint doi: https://doi.org/10.1101/2025.11.18.689171; this version posted November 19, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Table 1: Accession codes of the datasets used for predicting epigenomic profiles.

Assay	**GM12878**	**H1ESC**	**K562**
Hi-C	4DNFI1UEG1HD	4DNFIQYQWPF5	4DNFITUOMFUQ
DNase	ENCFF960FMM	ENCFF573NKX	ENCFF414OGC
CTCF	ENCFF644EEX	ENCFF332TNJ	ENCFF534ITO
H3K27ac	ENCFF469WVA	ENCFF919FBG	ENCFF381NDD
H3K27me3	ENCFF486WAD	ENCFF321JUJ	ENCFF242ENK
H3K4me3	ENCFF280PUF	ENCFF493QWY	ENCFF911JVK

Supplementary Table 2: Accession codes of the RNA-seq datasets used for validating cell type-specific motifs.

Assay	**GM12878**	**H1ESC**	**K562**
RNA-seq	ENCFF345SHY	ENCFF910OBU	ENCFF928NYA