Hi-Cformer enables multi-scale chromatin contact map modeling for single-cell Hi-C data analysis

Xiaoqing Wu[1,†], Xiaoyang Chen[1,†], Rui Jiang[1,*]

[1] Ministry of Education Key Laboratory of Bioinformatics, Bioinformatics Division at the Beijing National Research Center for Information Science and Technology, Center for Synthetic and Systems Biology, Department of Automation, Tsinghua University, Beijing 100084, China

[†] These authors contributed equally to this work.

[*] To whom correspondence should be addressed. Email: ruijiang@tsinghua.edu.cn

Abstract

Single-cell Hi-C captures the three-dimensional organization of chromatin in individual cells and provides insights into fundamental genomic processes such as gene regulation and transcription. While analyses of bulk Hi-C data have revealed multi-scale chromatin structures like A/B compartments and topologically associating domains, single-cell Hi-C data remain challenging to analyze due to sparsity and uneven distribution of chromatin contacts across genomic distances. These characteristics lead to strong signals near the diagonal and complex multi-scale local patterns in single-cell contact maps. Here, we propose Hi-Cformer, a transformer-based method that simultaneously models multi-scale blocks of chromatin contact maps and incorporates a specially designed attention mechanism to capture the dependencies between chromatin interactions across genomic regions and scales, enabling the integration of both global and fine-grained chromatin interaction features. Building on this architecture, Hi-Cformer robustly derives low-dimensional representations of cells from single-cell Hi-C data, achieving clearer separation of cell types compared to existing methods. Hi-Cformer can also accurately impute chromatin interaction signals associated with cellular heterogeneity, including 3D genome features such as topologically associating domain-like boundaries and A/B compartments. Furthermore, by leveraging its learned embeddings, Hi-Cformer can be extended to cell type annotation, achieving high accuracy and robustness across both intra- and inter-dataset scenarios.

Introduction

The 3D organization of chromatin plays a vital role in regulating genome function[1]. By shaping the spatial proximity between regulatory elements and genomic loci, 3D genome structure facilitates transcriptional regulation[2]. Moreover, the 3D organization of chromatin is closely associated with DNA replication timing[3,4], and contributes to the functional compartmentalization of the nucleus[5]. Disruptions in 3D genome structure have been shown to rewire regulatory architectures and cause pathogenic gene expression [6–8], underscoring the functional relevance of chromatin organization in the cellular context.

Chromosome conformation capture (3C) techniques[9], such as Hi-C[5,10] have significantly advanced the study of chromatin conformation, enabling the identification of 3D genome structures such as A/B compartments[5], topologically associating domains (TADs)[11–13], and chromatin loops[14]. In recent years, single-cell Hi-C (scHi-C) has enabled the investigation of 3D chromatin organization at the single-cell level, providing valuable insights into chromatin interactions that are associated with cell heterogeneity[15–19]. However, unlike other single-cell sequencing data that are typically represented as a feature vector per cell, scHi-C produces a series of matrices of sparse contact maps with pronounced genomic distance-dependent bias, leading to strong diagonal signals and complex multi-scale local patterns that pose challenges for downstream analysis[20].

Computational methods have been developed to address the challenges posed by scHi-C data and facilitate its analysis[21,22]. Non-deep learning methods typically adopt graph-based or statistical models as analytical frameworks[23–26]. For instance, random walk with restart (RWR) and its variants are widely used for imputing contact maps, with scHiCluster[23] being a representative example. Statistical methods such as principal component analysis (PCA) and latent Dirichlet allocation (LDA) are applied to obtain low-dimensional representations of cells[23,24]. Recent advances in deep learning have enabled the development of methods that capture the associations between complex cellular heterogeneity and chromatin interactions[27–30]. For example, Higashi[27], a representative self-supervised method, used hypergraph representation learning for the embedding and imputation of scHi-C data. scVI-3D[28] employed variational autoencoders to perform embedding and denoising of scHi-C data. scDEC-Hi-C[29] introduced a generative adversarial network (GAN)-based framework. scHiClassifier[30], a supervised method, was also developed for annotating cell types with scHi-C data.

Nevertheless, most existing methods model either entire chromosomal contact maps or pairwise interactions at a fixed resolution, without explicitly capturing the complex multi-scale structures present in chromatin contact matrices—structured patterns of varying sizes that carry cellular heterogeneity information about 3D genome organization[21]. This limitation hinders the simultaneous analysis of global and multi-scale local chromatin interaction features, restricting the ability of these methods to capture comprehensive chromatin interaction patterns.

To address these challenges, we propose Hi-Cformer, a transformer-based method that represents chromatin contact maps of a cell as sequences of token-like embeddings, analogous to the way language models encode sequences of words. The model integrates local structural features derived from multi-scale contact map blocks with global contextual information from entire chromosomal maps. Hi-Cformer employs a chromosome-aware hierarchical attention mechanism that captures intra-chromosomal dependencies of multi-scale interaction and enables effective

integration of inter-chromosomal context. This architecture preserves both fine-grained and overall chromatin interaction patterns, enabling comprehensive modeling of 3D genome organization similar to how language models capture contextual relationships within sequences.

The above modeling design of Hi-Cformer supports its strong performance on multiple downstream tasks. Hi-Cformer can generate informative cell representations from sparse and complex scHi-C data, capturing cellular heterogeneity and enabling accurate cell type discrimination across datasets with varying sequencing protocols, depths, and cell type compositions. Beyond dimensionality reduction, Hi-Cformer achieves more reliable imputation of sparse scHi-C contact maps than existing methods, by capturing cell-type-specific chromatin patterns while retaining single-cell-level heterogeneity. Hi-Cformer enhances the detection of cell-type-specific chromatin features, such as TAD-like domain boundaries and A/B compartments, and facilitates the analysis of cell-type heterogeneity in chromatin organization. As a flexible and extensible framework, Hi-Cformer can be adapted for supervised cell type annotation, showing strong performance in both intra- and inter-dataset scenarios. Together, these capabilities position Hi-Cformer as an effective and versatile tool for single-cell 3D genome analysis.

Results

Overview of Hi-Cformer

Hi-Cformer takes the intra-chromosomal contact maps of all chromosomes from a single cell at a given resolution as input, and outputs reconstructed contact maps at three hierarchical scales: a cell-level map formed by concatenating all intra-chromosomal maps, chromosome-level maps for individual chromosomes, and block-level submaps with varying resolutions. To achieve this objective, Hi-Cformer represents all intra-chromosomal contact maps of a cell as an ordered sequence of embeddings, following the input representation paradigm used in language models, thereby enabling the incorporation of both block-level and global-context 3D chromatin features while retaining cell-level resolution. Hi-Cformer consists of three modules: the multi-scale encoder module, the transformer module, and the multi-scale decoder module (Fig. 1a-1c).

The multi-scale encoder module encodes diagonal blocks of varying sizes from the intra-chromosomal contact map of each chromosome into fixed-dimensional vectors, referred to as block embeddings, as these regions correspond to genomically proximal loci with denser and more informative interaction patterns, often reflecting domain-like chromatin structures such as TADs. These block embeddings, together with chromosomal map embeddings of all chromosomes obtained via PCA, are then organized into a sequence ordered by chromosome index, block size, and position. Although not derived from a discrete token vocabulary, the resulting sequence plays a role analogous to token sequences in language models, enabling sequential modeling of chromatin contacts using a transformer.

The transformer module processes the output sequence from the multi-scale encoder, with additional embeddings for chromosome identity, block size, and genomic position added element-wise to each embedding in the sequence. These additional embeddings help the model distinguish among chromatin regions from different chromosomes, genomic positions, and scales. The resulting sequence is processed through transformer blocks[31] with modified attention masks that constrain block-level attention within chromosomes, as well as enabling global context integration through chromosomal map embeddings.

The multi-scale decoder module uses contextual block embeddings and chromosomal map embeddings derived from the output of the transformer module as input. These two types of embeddings are then fed into different multilayer perceptrons (MLPs) for three reconstruction tasks. The decoders produce cell embeddings as intermediate representations, along with multi-scale imputed contact maps as the outputs of these tasks.

Hi-Cformer is trained in a self-supervised manner to reconstruct input signals using a tailored masked language modeling (MLM) task[32], which encourages the model to capture long-range dependencies. The flexible architecture of Hi-Cformer allows introducing new embeddings to impute contact maps for chromatin regions of different resolutions, while also supporting supervised tasks like cell type annotation through the incorporation of a discriminator.

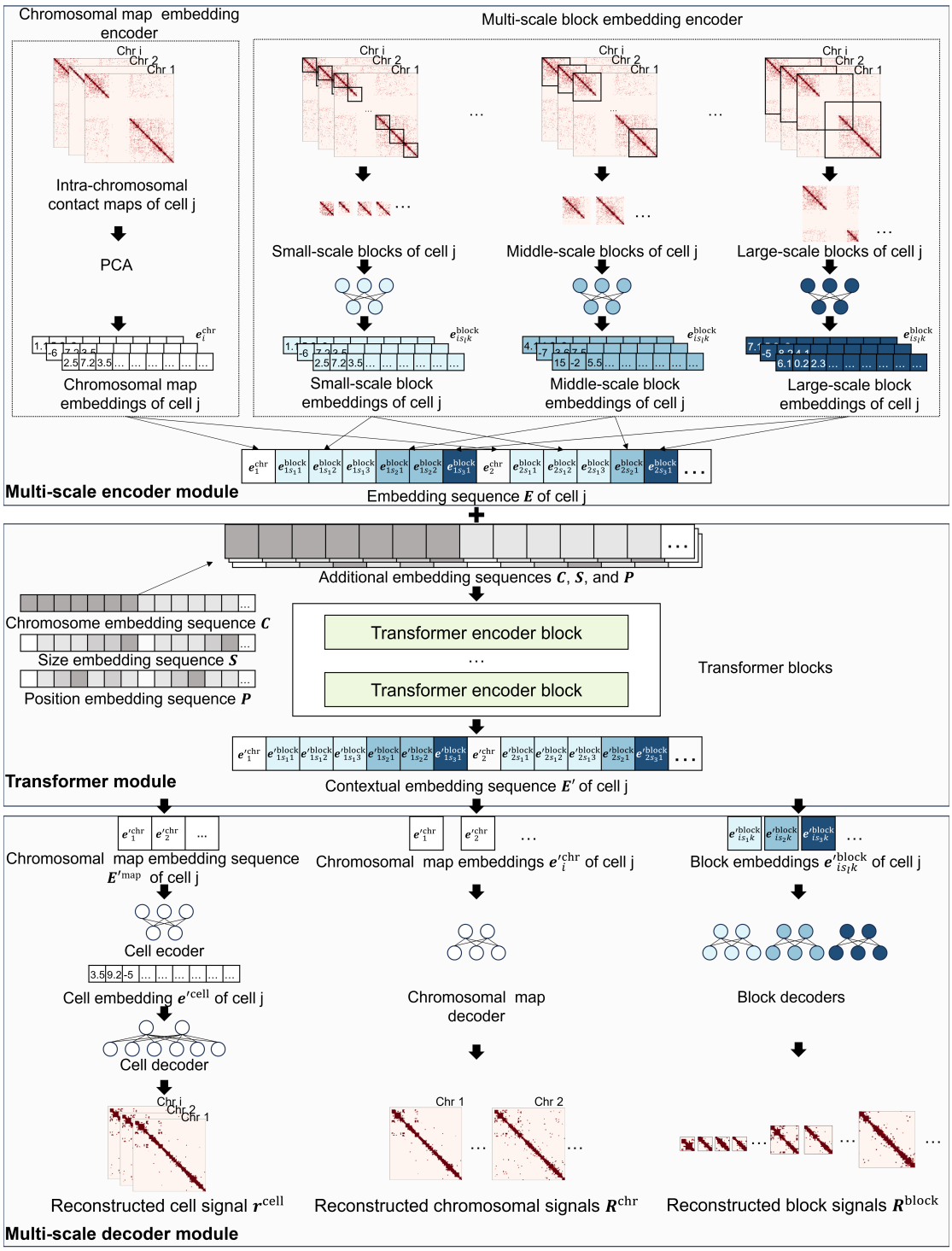


Fig. 1 | Overview of Hi-Cformer. a, The multi-scale encoder module embeds both intra-chromosomal contact maps and blocks at multiple scales using PCA and convolutional neural networks (CNNs). All embeddings are concatenated into an embedding sequence for each cell. **b**, The transformer module takes an embedding sequence as input and outputs a contextual embedding sequence. The embedding sequence is added to chromosome, size, and positional embeddings, then processed by transformer encoder blocks to capture intra- and inter-chromosomal dependencies. **c**, The multi-scale decoder module reconstructs signals at three levels: the cell-level contact map formed by concatenating all intra-chromosomal contact maps, chromosome-level contact maps for individual chromosomes, and block-level contact map blocks of varying size, using dedicated decoders for each level.

Hi-Cformer generates embeddings capturing rich cellular heterogeneity

Given the complex nature of raw scHi-C data, direct analysis of contact maps remains challenging. Therefore, obtaining low-dimensional representations of cells (referred to as cell embeddings) is essential for effective downstream analysis. We assessed the cell embeddings generated by Hi-Cformer on five scHi-C datasets[15–18] with known cell type labels at 1-Mb resolution, and compared Hi-Cformer against six baseline methods: Higashi[27], scDEC-Hi-C[29], HiCRep/MDS[26,33], scHiCluster[23], PCA[34], and LDA[24,35]. To quantitatively evaluate the ability of each method for revealing cellular heterogeneity, we used the Leiden algorithm[36] to cluster the cell embeddings, with the number of clusters set to match the ground-truth labels by tuning the resolution parameter via binary search. We then used normalized mutual information (NMI) and adjusted Rand Index (ARI) scores as metrics to evaluate clustering performance, where higher values indicate a more accurate representation of cellular heterogeneity. Further details on the datasets, experimental setup, evaluation criteria, and baseline methods are provided in Methods.

As shown in the left parts of Fig. 2a-2e, the cell embeddings derived from Hi-Cformer outperform those from the baselines in the clustering task, with average improvements of 6.41% in NMI and 29.35% in ARI over the second-best method, Higashi. We observed that non-deep learning methods, such as LDA, obtain substantial performance fluctuations across datasets. In contrast, Hi-Cformer demonstrates more stable and consistent performance. The superior clustering performance of Hi-Cformer is due to its ability to generate cell embeddings that effectively separate cell types, even subtypes, which are difficult to resolve for baseline methods. For example, on the Lee2019 dataset derived from the mouse hippocampus, Hi-Cformer separates closely related neuronal subtypes (Fig. 2b), which could not be identified using scHi-C data alone in the original study[16]. Hi-Cformer can also effectively distinguish cell types in datasets with small size and imbalanced populations. As illustrated in Fig. 2a, the embeddings generated by Hi-Cformer enable clear separation of the rare GM12878 cell type, which comprises only 33 cells, from other cell types, while baseline methods cannot obtain comparable results. All the results above highlight the superior capability of Hi-Cformer to capture cellular heterogeneity across diverse datasets.

During the sequencing procedure, dropout events often occur, leading to high noise and sparsity in the scHi-C data. To assess the robustness of Hi-Cformer under such conditions, we tested its performance on the simulated dataset by randomly removing raw reads from Ramani2017 dataset. Across all five datasets with various dropout rates, Hi-Cformer consistently outperforms baseline methods, demonstrating its strong robustness to dropout events (Fig. 2f). To investigate the contributions of each component and training strategy in Hi-Cformer, we performed ablation experiments by omitting the transformer module, preheating strategy, and MLM, respectively. The results show that removing the transformer module leads to a greater decline in embedding performance than removing preheating, as the transformer is essential for capturing intra- and inter-chromosomal dependencies of contact maps, while preheating mainly accelerates the training process. Notably, the models without the MLM task and those without the transformer module perform similarly, underscoring the critical role of the MLM task in effectively training the transformer module (Fig. 2g). Additionally, we explored the impact of selecting blocks of different sizes on embedding results. The results indicate that leveraging multi-scale blocks as input enhances embedding quality relative to single-scale designs (Supplementary Fig. S1). Together, these findings

emphasize the critical role of each component in enabling Hi-Cformer to effectively model the 3D chromatin organization.

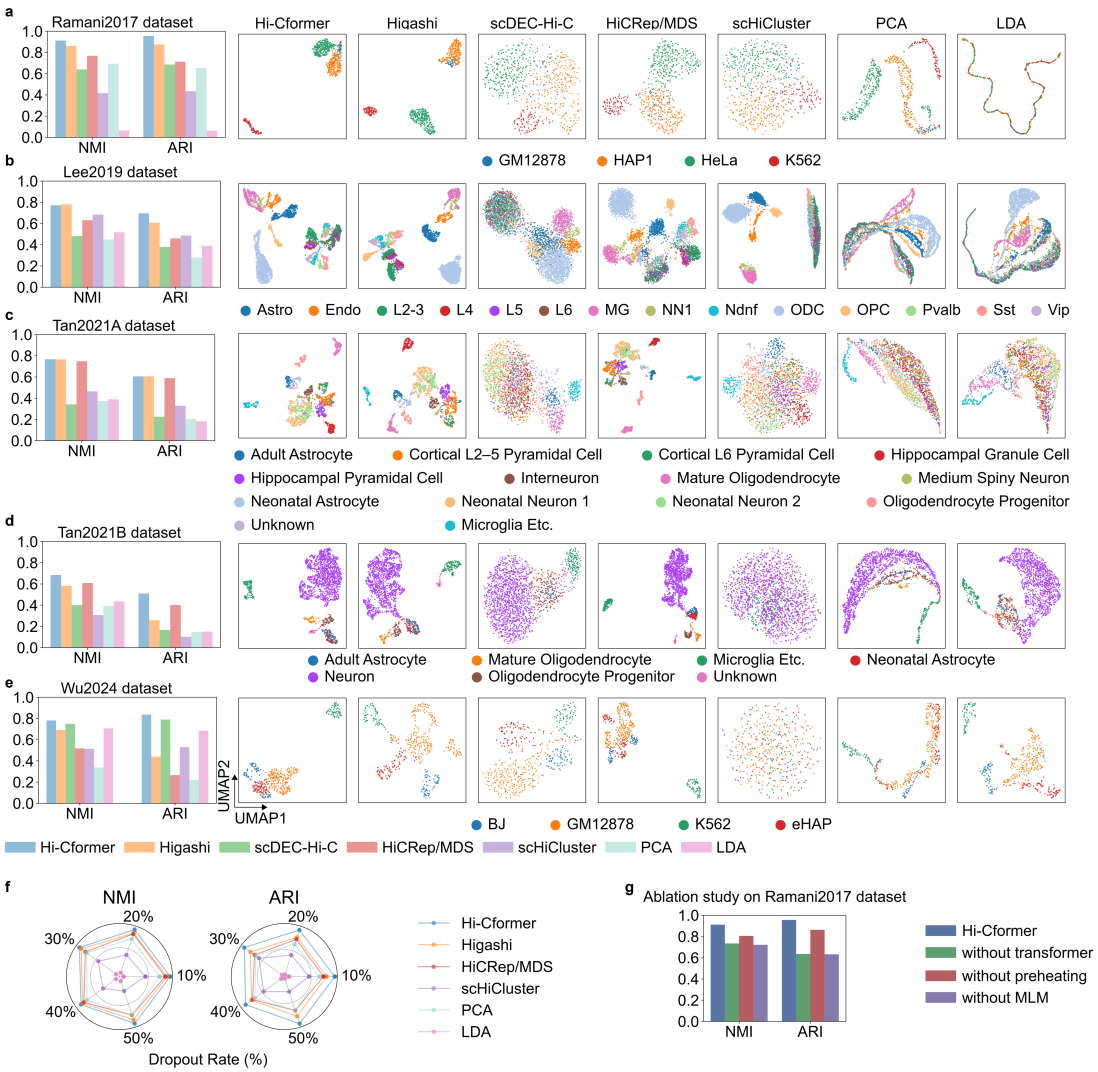


Fig. 2 | Evaluation and visualization of embeddings generated by different methods on scHi-C data. a–e, Quantitative evaluation of Hi-Cformer on the Ramani2017 (**a**), Lee2019 (**b**), Tan2021A (**c**), Tan2021B (**d**), and Wu2024 (**e**) datasets by comparing with six baseline methods: Higashi, scDEC-Hi-C, HiCRep/MDS, scHiCluster, PCA, and LDA. Cell embeddings were generated from contact maps at 1-Mb resolution and clustered using the Leiden algorithm with the number of clusters set to match the ground-truth labels. Left panels report clustering performance measured by normalized mutual information (NMI) and adjusted Rand index (ARI); right panels report UMAP visualization right panels report UMAP visualization of the embeddings, colored by ground-truth cell types. **f**, Robustness analysis on the Ramani2017 dataset under simulated dropout noise. **g**, Ablation analysis of model components on the Ramani2017 dataset. Comparison of performance under different ablation settings, i.e., omitting the transformer module, the preheating strategy, or the MLM task.

Hi-Cformer provides reliable imputation of scHi-C contact maps

Due to technical limitations during sequencing, scHi-C data typically have the characteristics of sparsity and low signal-to-noise ratio, prompting the use of imputation to improve data quality[21]. Using scHi-C data from the Ramani2017 dataset at 1-Mb resolution, we evaluated Hi-Cformer against two imputation methods, Higashi and scHiCluster. Compared to bulk sequencing data derived from the same cell types in the Rao2014 dataset[14], contact maps imputed by Hi-Cformer exhibit a median improvement of 68.95% in Pearson correlation coefficients (PCCs) and 66.15% in cosine similarities over the second-best method, Higashi, averaged across two cell types (Fig. 3a). Additionally, using pseudo-bulk contact maps averaged from single-cell data of the same cell type as the reference, Hi-Cformer achieves median improvements of 70.60% in PCCs and 68.15% in cosine similarities over Higashi, averaged across two cell types (Fig. 3b). The above results demonstrate the ability of Hi-Cformer to accurately enhance cellular heterogeneity and filter out irrelevant noise when imputing signals.

Furthermore, to enable a quantitative evaluation of imputation accuracy at the level of individual cells, we benchmarked Hi-Cformer and baselines on five simulated datasets by introducing varying degrees of dropout events to the raw scHi-C contact maps. The raw contact maps before dropouts serve as ground truth, allowing for direct comparison of the ability of each method to reconstruct missing chromatin interactions. Hi-Cformer consistently outperforms Higashi and scHiCluster across all four evaluation metrics, including PCC, cosine similarity, structural similarity index measure (SSIM), and peak signal-to-noise ratio (PSNR), and also maintains stable performance across varying dropout rates, indicating high accuracy and robustness for imputation (Fig. 3c-f). The lower concordance with ground truth obtained by Higashi and scHiCluster may be partially related to their tendency to generate denser contact maps during imputation.

Finally, to assess whether imputation with Hi-Cformer can improve accuracy of downstream analyses, we applied PCA on contact maps imputed by different methods to reduce dimensionality to 64 components, followed by benchmarking the performance of cell clustering using corresponding metrics. The results demonstrated that imputation with Hi-Cformer can provide a more consistent and stable improvement in cell clustering than that with baseline, suggesting the potential for effectively uncovering the cell-type heterogeneity in scHi-C data (Fig. 3d).

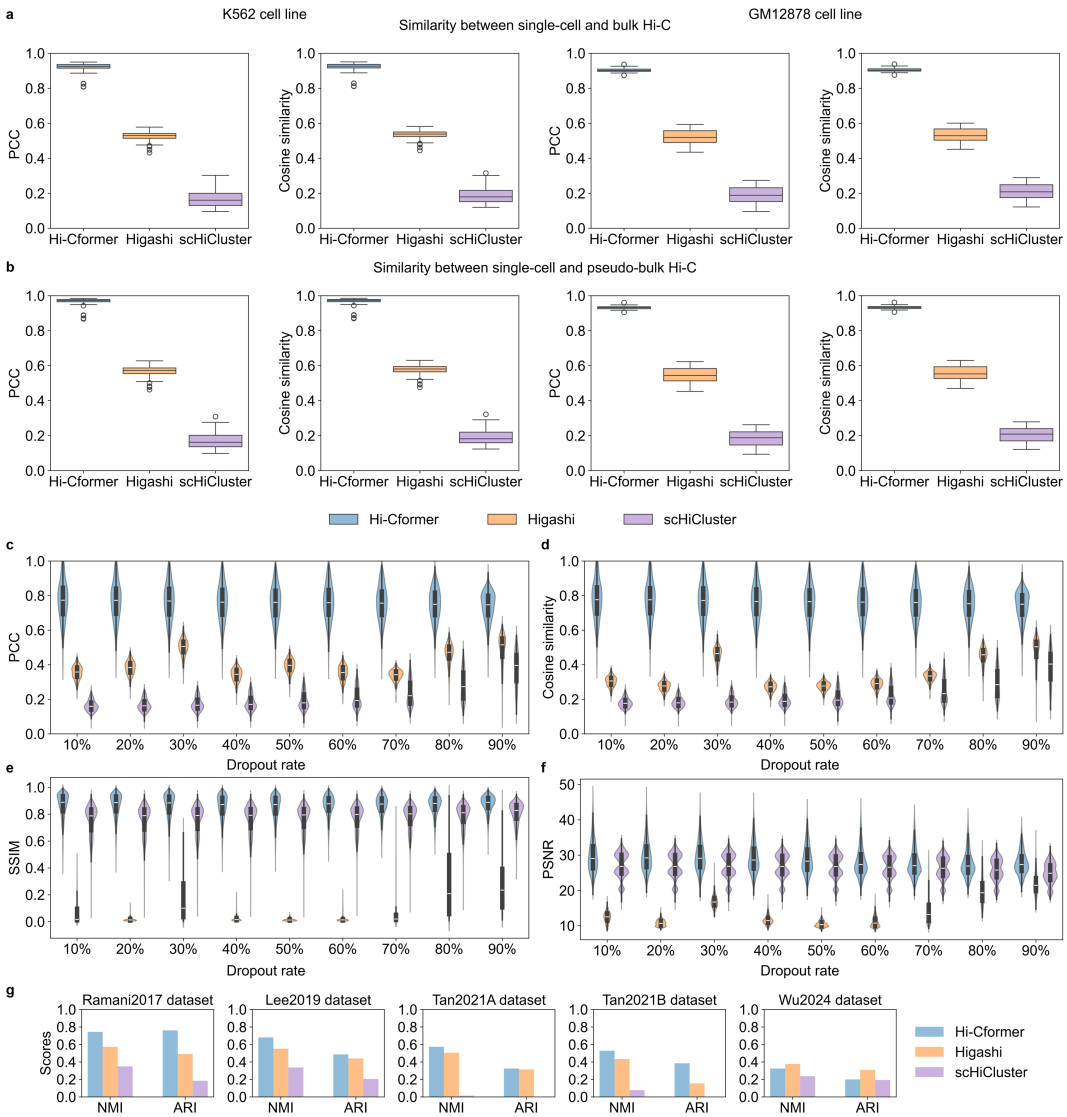


Fig. 3 | Comparison of Hi-Cformer with two baseline imputation methods from different evaluation perspectives. a, Pearson correlation coefficients (PCCs) and cosine similarities between scHi-C data imputed by three methods from the Ramani2017 dataset and corresponding bulk Hi-C data from Rao2014 dataset, for the K562 and GM12878 cell types (the left two subpanels in a correspond to K562, and the right two to GM12878). **b**, PCCs and cosine similarities between scHi-C data imputed by three methods from the Ramani2017 dataset and corresponding pseudo-bulk contact maps (aggregated from raw single-cell maps), for the K562 and GM12878 cell types (the left two subpanels in b correspond to K562, and the right two to GM12878). **c**, PCCs between imputed and ground truth data on simulated dropout datasets generated from the Ramani2017 dataset with varying dropout rates (10% to 90%). **d**, Cosine similarities between imputed and ground truth data on simulated dropout datasets generated from the Ramani2017 dataset with varying dropout rates (10%-90%). **e**, SSIMs between imputed and ground truth data on simulated dropout datasets generated from the Ramani2017 dataset with varying dropout rates (10% to 90%). **f**, PSNRs between imputed and ground truth data on simulated dropout datasets generated from the Ramani2017 dataset with varying dropout rates (10% to 90%). In **a-f**, the upper, middle, and lower edges of the boxes denote the upper quartile, median, and lower quartile, respectively, and error bars denote the maximum and minimum values. **g**, NMI and ARI scores for clustering results after dimensionality reduction of imputed data by three methods on five benchmark scHi-C datasets.

Hi-Cformer facilitates the identification of cell-type specific structures

Cell-type-specific structures of chromatin contact maps have been a central focus of scHi-C analysis[37]. In this section, we explored how imputed scHi-C data with Hi-Cformer facilitates the identification of cell-type specific structures such as TAD-like domain boundaries and A/B compartments.

We first focused on a genomic region (chr9:132,950,000-138,000,000) previously reported to differ between the K562 and GM12878 cell types[29]. Within this region, the ABL1 gene locus (chr9:133,710,641-133,763,062) is of particular interest. ABL1 is a known oncogene involved in chronic myeloid leukemia (CML). While both BCR and ABL1 are present in K562 and GM12878 cells, only K562 harbors the BCR-ABL fusion gene, resulting from a translocation between chromosome 9 and 22[38]. This may contribute to the differences observed in the scHi-C contact maps of this region. To better resolve this region, we leveraged Hi-Cformer's ability to encode contact map blocks of varying sizes into fixed-dimensional embeddings, which enables flexible resolution handling. Specifically, we introduced an additional embedding for the selected 50-kb resolution region, accompanied by a dedicated encoder and decoder, enabling Hi-Cformer to impute this high-resolution region (details in Methods). We conducted imputation experiments for this region on K562 and GM12878 cells from the Ramani2017 dataset and presented the resulting chromatin contact maps. For both cell types, raw and imputed scHi-C contact maps were averaged to generate pseudo-bulk data, which were then compared to the corresponding bulk Hi-C maps from Rao2014 dataset[14]. As shown in Fig. 4a, the imputed pseudo-bulk maps more clearly revealed structural differences (chr9:133,550,000-134,050,000) between K562 and GM12878 cells, particularly around the ABL1 locus. Within this genomic region, H3K27ac ChIP-seq signals are enriched in K562 cells compared to GM12878 cells (Fig. 4b), consistent with enhanced regulatory activity associated with the oncogenic fusion event. In addition to directly inspecting local structural differences, we further assessed large-scale chromatin organization by analyzing the merged scHi-C correlation matrices. Imputation with Hi-Cformer strengthened the checkerboard-like patterns indicative of A/B compartmentalization (Fig. 4c).

We further explored the impact of imputation on the identification of structural features in scHi-C data, including TAD-like domains and A/B compartments. TADs are regions of high chromatin contact frequency observed in ensemble Hi-C contact maps[11]. Recent studies have reported similar structures, TAD-like domains, in scHi-C contact maps[39]. We extracted scHi-C contact maps for the chr9: 132,950,000-138,000,000 at 50-kb resolution, both before and after imputation, from the Ramani2017 dataset. We then performed TAD-like domain boundaries calling using the method in Higashi[27] that identifies local minima of single-cell insulation scores as domain boundaries. The imputed data reveal clearer TAD-like domain boundaries than the raw data (Fig. 4d and Supplementary Fig. S2). By organizing single-cell insulation scores by cell type, we observed certain TAD-like domain boundaries that appear to be cell type-specific. For instance, in the region highlighted by the red box in Fig. 4d, the single-cell insulation scores of HeLa cells display clearer TAD-like boundaries relative to those from other cell types. A/B compartments reflect the segregation of the genome into transcriptionally active and inactive regions at the mega-base scale[5]. We sought to explore how imputation influences the identification of A/B compartments. We extracted scHi-C contact maps of the chr1 at 1-Mb resolution, both before and after imputation,

from the two cell types (K562 and GM12878) in the Ramani2017 dataset. Then we conducted A/B compartment calling using the method in Higashi[27], which generates comparable continuous A/B compartment scores by applying PCA to the pooled Pearson correlation matrix and projecting individual single-cell matrices using the resulting PCA components. Compared with the A/B compartment annotations derived from raw single-cell data, the compartment scores obtained from imputed data are more consistent with those from pooled pseudo-bulk data. To assess the biological relevance of the compartment scores, we examined their correspondence with H3K27ac ChIP-seq signals[40]. Regions enriched for H3K27ac tend to display scores with consistent polarity—either predominantly positive or negative—supporting the view that the scores reflect meaningful patterns of chromatin activity and compartmentalization (Fig. 4e and Supplementary Fig. S3).

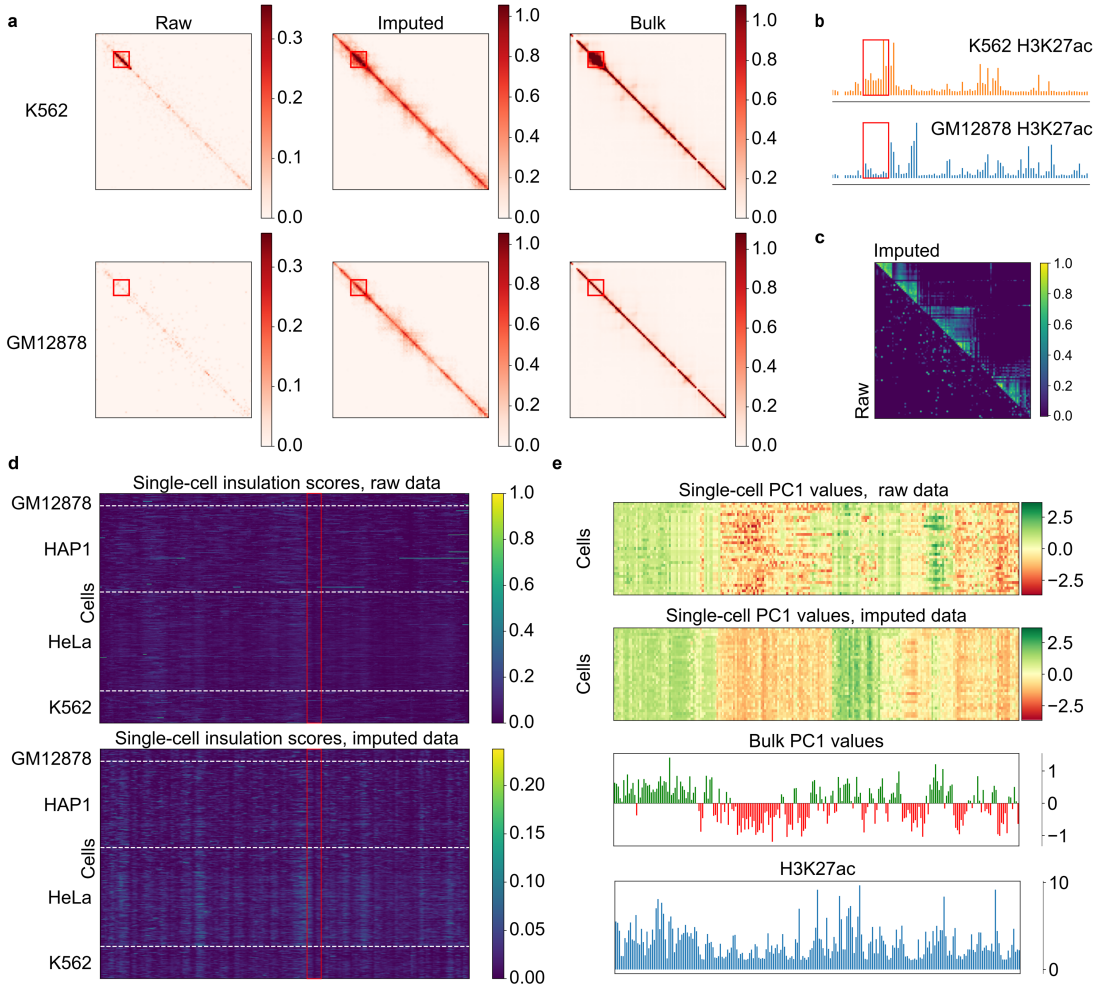


Fig. 4 | Multi-scale chromatin structure analysis from raw and imputed scHi-C data using Hi-Cformer. a, Averaged contact maps at 50-kb resolution for K562 and GM12878 cells (top and bottom rows) in the chr9:132,950,000-138,000,000, shown as raw single-cell data (left), imputed data using Hi-Cformer (middle), and corresponding bulk Hi-C data (right). Red boxes highlight a subregion (chr9:133,550,000-134,050,000) around the ABL1 locus. **b**, ChIP-seq tracks of H3K27ac in GM12878 and K562 across the same region as in **a**, with the red box indicating the same subregion around the ABL1 locus. **c**, Pearson correlation matrix computed from pseudo-bulk contact maps (chr9:132,950,000-138,000,000, 50-kb resolution) of GM12878 cells, obtained by averaging single-cell contact maps from Ramani2017 dataset. Each entry reflects the Pearson correlation between rows of the contact

map. Lower and upper triangles show results before and after imputation, respectively. **d**, Single-cell insulation scores in chr9:100,000,000-140,000,000 at 50-kb resolution for four cell types (GM12878, HAP1, HeLa, and K562), computed from raw (top) and imputed (bottom) data. Red boxes indicate example TAD-like boundaries. **e**, Single-cell values of the first principal component (PC1) representing A/B compartment scores for GM12878 of chr1 at 1-Mb resolution. From top to bottom: PC1 values derived from raw single-cell data, PC1 values from imputed single-cell data, bulk-level PC1 values, and H3K27ac ChIP-seq signals. All tracks are shown over the same genomic region at the same resolution for direct comparison.

Supervised cell type annotation on scHi-C data with Hi-Cformer

As a versatile framework, Hi-Cformer can also be adapted for supervised cell type annotation, leveraging the existing labeled scHi-C data to directly identify cell types without relying on unsupervised clustering, thus avoiding the cumbersome process of manual annotation. This adaptation involves incorporating a simple discriminator on top of the low-dimensional cell embeddings, with the addition of a classification loss function during training. We compared Hi-Cformer with three supervised cell type annotation methods from Zhou et al.[30]: scHiClassifier, logistic regression (LR), and random forest (RF). For all baselines, we followed the same feature extraction and training procedures as described in their original study. Evaluation was first conducted on five scHi-C datasets using five-fold cross-validation. As shown in Fig. 5a, Hi-Cformer consistently outperforms the other methods, particularly on datasets (Lee2019 dataset and Tan2021A dataset) with complex and closely related cell types. Hi-Cformer also demonstrates superior performance in identifying rare cell types. In the Tan2021A dataset, the medium spiny neuron, a rare cell type comprising only 37 cells, is effectively recognized by Hi-Cformer. In contrast, RF fails to annotate this cell type, while scHiClassifier misclassifies many other cell types as medium spiny neurons (Fig. 5c). These results highlight the effectiveness of Hi-Cformer in intra-dataset classification tasks. To further assess these methods under more practical and challenging conditions, we evaluated their performance in inter-dataset cell type annotation, where batch effects such as technical variations may affect annotation accuracy. Tan2021A dataset consists of cells from the cortex and the hippocampus across the first postnatal year, while Tan2021B dataset contains the visual cortex of dark-reared and control mice during the first month. The biological and experimental differences between these datasets pose significant challenges for inter-dataset evaluation. After unifying the label granularity between the two datasets (details in Methods), we trained Hi-Cformer and baseline methods on one dataset and evaluated them on the other. These results indicate that Hi-Cformer maintains superior performance in inter-dataset annotation (Fig. 5d), demonstrating its adaptability across datasets.

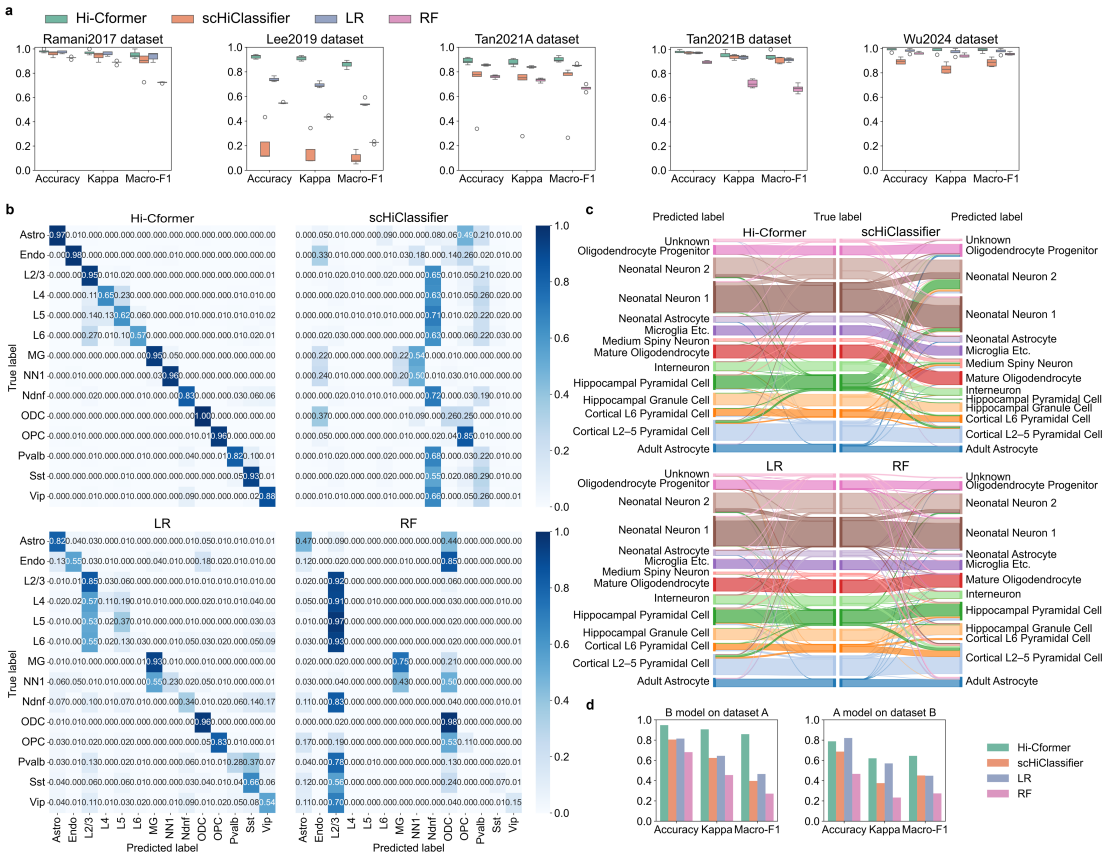


Fig. 5 | Comparison of Hi-Cformer with baseline methods for supervised cell type annotation on scHi-C data. Hi-Cformer, scHiClassifier, logistic regression (LR), and random forest (RF) were evaluated for cell type classification performance across five public scHi-C datasets. **a**, Five-fold cross-validation results on the Ramani2017, Lee2019, Tan2021A, Tan2021B, and Wu2024 datasets are shown,, evaluated using accuracy, Cohen's kappa, and macro-F1 score. The upper, middle, and lower edges of the boxes denote the upper quartile, median, and lower quartile, respectively, and error bars denote the maximum and minimum values, excluding outliers. **b**, Confusion matrix derived from different methods on the Lee2019 dataset. The results are aggregated from all validation folds in a five-fold cross-validation. Each value in the matrix represents the fraction of samples predicted for a given cell type among all samples of the true cell type. **c**, Sankey plot for illustrating annotated results on the Tan2021A dataset for each cell type. Ground truth labels are shown in the center, and predicted cell types are shown on the left and right. **d**, Inter-dataset annotation performance of Hi-Cformer and baseline methods between the Tan2021A and Tan2021B datasets, evaluated using accuracy, Cohen's kappa, and macro-F1 score.

Discussion

In this study, we present Hi-Cformer, a transformer-based framework for analyzing scHi-C data. Hi-Cformer generates embeddings that effectively preserve cell type consistency, facilitating clearer distinctions among diverse cell populations. Its imputation robustly recovers missing contacts in scHi-C data, improving data quality and revealing cell type-specific regions. The variant of Hi-Cformer leverages scHi-C features for accurate cell type annotation, underscoring its potential for comprehensive single-cell chromatin conformation analysis.

The key innovation of Hi-Cformer lies in its multi-scale transformer architecture, which

models chromatin contact patterns across different scales while maintaining single-cell resolution. The model captures both fine-grained and overall chromatin interaction patterns within individual cells. The chromosome-aware attention mechanism enables Hi-Cformer to capture both intra- and inter-chromosomal dependencies, integrating long-range structural relationships in a data-driven manner without relying on predefined neighborhoods or fixed-scale assumptions. Furthermore, the design of Hi-Cformer offers practical flexibility, permitting straightforward modifications to incorporate analysis of regions of interest at different resolutions and to support downstream tasks such as cell type annotation with relative ease.

There are several potential directions to further improve Hi-Cformer. First, as a transformer-based model for scHi-C data, Hi-Cformer represents input contact maps as sequences of token-like embeddings. This token-based formulation, along with additional embeddings, provides a flexible basis for incorporating other omics modalities. For example, DNA methylation or gene expression data could be introduced through newly designed embeddings, allowing joint modeling in a unified framework[16,41]. Second, potential extensions of Hi-Cformer also include the integration of inter-chromosomal contacts and the modeling of higher-order multi-way interactions to more comprehensively capture the complexity of 3D genome organization. Incorporating inter-chromosomal contacts enables representation of spatial relationships across chromosomes, which are frequently absent in current models. The inclusion of multi-way interactions facilitates characterization of chromatin structures that extend beyond pairwise contacts. These extensions may be implemented within the Hi-Cformer framework through the development of new encoding and decoding modules. Collectively, such advancements have the potential to improve the analysis of scHi-C data, offering new opportunities for advancing 3D genome research. Finally, Hi-Cformer, as a transformer-based model trained with a self-supervised objective using MLM, aligns with recent trends in foundation models, offering a scalable framework for learning general-purpose representations of 3D genome organization. With sufficient training on large and diverse datasets, such models could support transfer learning across biological conditions and tasks, enabling various downstream applications.

Methods

Model architecture of Hi-Cformer

Multi-scale encoder module. Hi-Cformer takes as input a collection of intra-chromosomal contact maps from a single cell, each corresponding to a specific chromosome and generated at the same resolution. Due to the varying lengths of individual chromosomes, these contact maps differ in shape. Consequently, they are organized as a list of matrices and fed into the multi-scale encoder module. The multi-scale encoder module comprises two components: the chromosomal map embedding encoder and the multi-scale block embedding encoder. The former is designed to capture global chromatin structure by flattening all contact maps corresponding to the same chromosome in the training set, followed by PCA to obtain fixed-dimensional embeddings. These embeddings are referred to as the chromosomal map embeddings, denoted as $\boldsymbol{e}_i^{\mathrm{chr}} \in \mathbb{R}^d$. The multi-scale block embedding encoder captures block-level features through a set of convolutional neural networks (CNNs) with varying kernel sizes (s_i = 8, 16, 32, 64, or 128), where each CNN has an input channel of 1 and an output channel of dimension d. Each CNN employs a stride equal to its kernel width. Unlike conventional CNNs that scan the intra-chromosomal map, these networks are specifically designed to process square blocks of varying sizes along the diagonal of the contact map, thereby focusing on regions enriched with chromatin contacts across multiple genomic scales. This design implies that each convolutional kernel is applied with equal stride along both the row and column directions of a contact map. The multi-scale block embedding encoder embeds each diagonal block into a fixed-dimensional embedding, referred to as the block embedding $\boldsymbol{e}_{is_lk}^{\mathrm{block}} \in \mathbb{R}^d$ (denoting the k-th block embedding of chromosome i with size s_l). By default, the model operates on scHi-C contact maps at 1-Mb resolution, with the embedding dimension d = 128. After obtaining individual embeddings, we arrange them sequentially according to chromosome index, block size, and position: $\left[\boldsymbol{e}_1^{\mathrm{chr}}, \boldsymbol{e}_{1s_11}^{\mathrm{block}}, \boldsymbol{e}_{1s_12}^{\mathrm{block}}, \ldots, \boldsymbol{e}_{1s_21}^{\mathrm{block}}, \ldots, \boldsymbol{e}_2^{\mathrm{chr}}, \ldots\right]$ to form embedding sequence $\boldsymbol{E} \in \mathbb{R}^{n\times d}$, where n denotes the total number of embeddings. This ordered embedding sequence serves as the input to the transformer module, enabling Hi-Cformer to model chromatin context across chromosomes and scales in a manner analogous to language models processing token sequences.

Transformer module. Similar to the positional embedding in the natural language processing (NLP) domain, in order to form a sentence that can be input into the transformer encoder, the embedding sequence requires the addition of extra information embedding to distinguish between different embeddings.

We designed three types of trainable information embedding: (1) Chromosome embedding, used to distinguish embeddings from different chromosomes, $\{\boldsymbol{c}_i \in \mathbb{R}^d \mid i = 1, 2, \ldots, c\}$, where d is the dimension of embedding and each embedding including chromosomal map embedding is assigned an embedding corresponding to its chromosome and c represents the number of chromosomes in the scHi-C data. (2) Size embedding, used to distinguish block embeddings of different scales, $\{\boldsymbol{s}_i \in \mathbb{R}^d \mid i = 1,2,\ldots,s\}$, where s is the number of different scales, and each block embedding is assigned an embedding corresponding to its scale. (3) Position embedding, used to distinguish block embeddings within the same chromosome and scale, $\{\boldsymbol{p}_i \in \mathbb{R}^d \mid i = 1,2,\ldots,m\}$, where m is the number of block embeddings for the smallest scale in the largest chromatin contact map, representing the maximum number of block embeddings within the same chromosome and scale. The information embeddings are added to the original embedding

sequence, resulting in new embedding sequence $\widetilde{\boldsymbol{E}}$.

$$\tilde{\boldsymbol{e}}_{i,s_l,k}^{\text{block}} = \boldsymbol{e}_{i,s_l,k}^{block} + \boldsymbol{c}_i + \boldsymbol{s}_l + \boldsymbol{p}_k$$

$$\tilde{\boldsymbol{e}}_i^{\text{chr}} = \boldsymbol{e}_i^{\text{chr}} + \boldsymbol{c}_i$$

As a module of the transformer-based model Hi-Cformer, the transformer module by default consists of four stacked transformer encoder blocks with a modified attention mechanism. This mechanism can model the dependencies of chromosomal interactions at different positions, allowing it to capture the chromosomal interaction network within a cell. Each transformer block operates on inputs and outputs with an embedding dimension of d and utilizes 8 attention heads. The input and output of each stacked transformer block can be formulated as:

$$\widetilde{\boldsymbol{E}}^{(0)} = \widetilde{\boldsymbol{E}},$$

$$\widetilde{\boldsymbol{E}}^{(l)} = \text{transformer_encoder_block}\left(\widetilde{\boldsymbol{E}}^{(l-1)}\right)$$

where $\widetilde{\boldsymbol{E}}^{(l)}$ is the output of the l-th transformer encoder block, and $\text{transformer_encoder_block}(\cdot)$ is a single transformer block. The modified attention mechanism captures the interactions between embeddings and can be formulated as follows:

$$Q = \widetilde{\boldsymbol{E}}^{(l-1)} W_Q^{(l)},\ K = \widetilde{\boldsymbol{E}}^{(l-1)} W_K^{(l)},\ V = \widetilde{\boldsymbol{E}}^{(l-1)} W_V^{(l)},$$

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}} + \text{attn_mask}\right) V$$

$$\text{attn_mask}_{ij} = \begin{cases} 0, & \text{if the } j\text{-th embedding is within attention scope of the } i\text{-th embedding} \\ -\infty, & \text{otherwise} \end{cases}$$

where $W_Q^{(l)}$, $W_K^{(l)}$, and $W_V^{(l)}$ are trainable weight matrices used to compute the Query (Q), Key (K), and Value (V), respectively, and d_k denotes the dimension of the Key vector. The attn_mask$\in \mathbb{R}^{n\times n}$ is a matrix whose elements are either 0 or negative infinity, with the values being customizable. Specifically, setting attn_mask_{ij} to negative infinity ensures that during the computation of attention scores, the i-th embedding will no longer attend to the j-th embedding, but will only compute attention scores with embeddings where $\text{attn_mask}_{ik} = 0$. By modifying the attn_mask matrix, the interaction relationships between embeddings can be better adjusted. By default, block embeddings can only compute attention with other embeddings within the same chromosome (including both chromosomal map embeddings and block embeddings), and chromosomal map embeddings, in addition to attending to block embeddings within the same chromosome, also attend to all other chromosomal map embeddings.

The embedding sequence $\widetilde{\boldsymbol{E}}$ is fed into the transformer module, and the resulting output is denoted as $\boldsymbol{E}'$. The contextual chromosomal map embeddings in $\boldsymbol{E}'$ after the transformer module can be seen as containing not only all the information of their respective chromosomes but also information from other chromosomes. And each block embedding not only contains information about its corresponding contact region but also includes additional information from other regions within the same chromosome.

Multi-scale decoder module. The multi-scale contact map decoder module reconstructs the original signal by passing different types of contextual embeddings through MLPs while also performing dimensionality reduction on all contextual chromosomal map embeddings of a cell to obtain the cell embedding. The cell-level component consists of a two-layer fully connected neural

network. The first layer takes as input a concatenation of all contextual chromosomal map embeddings $\boldsymbol{E}'^{\text{map}} \in \mathbb{R}^{c\times D}$ from $\boldsymbol{E}'$, and outputs the cell embedding $\boldsymbol{e}'^{\text{cell}} \in \mathbb{R}^{g}$ where $g = 64$, which is the specified low-dimensional cell representation. The second layer takes $\boldsymbol{e}'^{\text{cell}}$ as input and outputs $\boldsymbol{r}^{\text{cell}} \in \mathbb{R}^{o}$, whose dimension matches the flattened and concatenated upper triangular matrices of all chromatin contact frequency maps, considering their symmetry. From the output of the first layer, we obtain the cell embedding $\boldsymbol{e}'^{\text{cell}}$ while the output of the second layer provides the denoised chromatin contact frequency maps $\boldsymbol{r}^{\text{cell}}$. The chromosome-level component consists of a single-layer fully connected neural network. It takes a contextual chromosomal map embedding $\boldsymbol{e}'^{\text{chr}}_{i}$ as input and outputs a matrix $\boldsymbol{R}^{\text{chr}}_{i} \in \mathbb{R}^{f\times f}$, where f is set to 112 in this work, to reconstruct individual chromosomal contact frequency maps. The reconstructed signals from the chromosome-level decoder are considered scaled versions of the chromosomal contact frequency maps. Therefore, the original frequency maps are interpolated to the specified size before computing the reconstruction loss. The block-level decoder consists of several single-layer fully connected neural networks, taking an input $\boldsymbol{e}'^{\text{block}}_{i,s_l,k} \in \mathbb{R}^{d}$ and producing an output $\boldsymbol{R}^{\text{block}}_{i,s_l,k} \in \mathbb{R}^{s_l\times s_l}$ to reconstruct the corresponding block. The purpose of these decoders is to train the model through reconstruction tasks, thereby obtaining low-dimensional representations of multi-scale structures, including cells, chromosomes, and blocks.

Model training

The model training employs reconstruction tasks at three different scales, with reconstruction losses computed separately. The overall loss function is the weighted sum of the losses from these three reconstruction tasks. Assume that $\boldsymbol{r}^{\text{cell}}, \boldsymbol{R}^{\text{chr}}_{i}$, and $\boldsymbol{R}^{\text{block}}_{i,s_l,k}$ correspond to the original signals $\boldsymbol{g}^{\text{cell}}, \boldsymbol{G}^{\text{chr}}_{i}$, and $\boldsymbol{G}^{\text{block}}_{i,s_l,k}$. Then, the training loss function is calculated as follows:

$$\text{Loss}_{\text{cell}} = \frac{1}{o}\sum_{i=1}^{o}\left(\boldsymbol{g}^{\text{cell}}[i] - \boldsymbol{r}^{\text{cell}}[i]\right)^2$$

$$\text{Loss}_{\text{chr}} = \frac{1}{c\cdot f^2}\sum_{i=1}^{c}\sum_{j=1}^{f}\sum_{k=1}^{f}\left(\boldsymbol{G}^{\text{chr}}_{i}[j,k] - \boldsymbol{R}^{\text{chr}}_{i}[j,k]\right)^2$$

$$\text{Loss}_{\text{block}} = \frac{1}{\sum_{i=1}^{c}\sum_{s}(N_{is}\cdot p_i^2)} \times \sum_{i=1}^{c}\sum_{l=1}^{s}\sum_{k=1}^{N_{il}}\sum_{m=1}^{p_i}\sum_{n=1}^{p_i}\left(\boldsymbol{G}^{\text{block}}_{i,s_l,k}[m,n] - \boldsymbol{R}^{\text{block}}_{i,s_l,k}[m,n]\right)^2$$

$$\text{Loss} = \alpha\cdot\text{Loss}_{\text{cell}} + \beta\cdot\text{Loss}_{\text{chr}} + \gamma\cdot\text{Loss}_{\text{block}}$$

Here, α, β, and γ are tunable hyperparameters. Prior to the main training, Hi-Cformer undergoes a phase referred to as preheating. During preheating, the transformer module is removed, while other modules and loss functions remain unchanged. The only difference lies in the inputs to the decoder. Specifically, $\boldsymbol{E}'^{map}$, $\boldsymbol{e}'^{\text{chr}}_{i}$, and $\boldsymbol{e}'^{\text{block}}_{i,s_l,k}$ extracted from $\boldsymbol{E}'$ are replaced by $\widetilde{\boldsymbol{E}}^{map}$, $\tilde{\boldsymbol{e}}^{\text{chr}}_{i}$, and $\tilde{\boldsymbol{e}}^{\text{block}}_{i,s_l,k}$ extracted from $\widetilde{\boldsymbol{E}}$. The decoder then produces $\boldsymbol{r}'^{\text{cell}}, \boldsymbol{R}'^{\text{chr}}_{i}$, and $\boldsymbol{R}'^{\text{block}}_{i,s_l,k}$, which replace $\boldsymbol{r}^{\text{cell}}, \boldsymbol{R}^{\text{chr}}_{i}$, and $\boldsymbol{R}^{\text{block}}_{i,s_l,k}$ in the loss computation, allowing the calculation of the same losses as described above.

The application of MLM tasks in the BERT model[32] has shown remarkable efficiency and brought profound changes to the field of NLP. Inspired by these masked language modeling tasks, this work also designs a masked language modeling task to train the model. Specifically, given a

masking ratio, after obtaining the embedding sequence $\boldsymbol{E}$, all embeddings (including both chromosomal map and block embeddings) belonging to the same chromosome are randomly replaced with a learnable special embedding [MASK] at the specified ratio. The modified embedding sequence $\boldsymbol{E}$ is then used as the input to the transformer module. In our experiments, the default masking ratio is set to 20%. For highly sparse contact maps, we recommend increasing the masking ratio to encourage more effective learning of missing or noisy regions. For less sparse data, a lower masking ratio may be preferable to preserve more of the original structural information during training.

Data processing

In this work, we utilized several publicly available scHi-C datasets with cell type labels, as well as bulk Hi-C and ChIP-seq datasets for reference analyses. The scHi-C datasets include the Ramani2017 dataset[15] (GEO: GSE84920), Lee2019 dataset[16] (GEO: GSE130711), Tan2021A dataset[17] (GEO: GSE164203), Tan2021B dataset [17] (GEO: GSE146397), Wu2024 dataset[18] (GEO: GSE240128). For bulk Hi-C reference, we used contact maps for the K562 and GM12878 cell lines from the Rao2014 dataset[14] (GEO: GSE63525). In addition, we obtained ChIP-seq tracks of H3K27ac in GM12878 and K562 (GEO: GSE29611) generated by ENCODE[40] to support downstream validation and biological interpretation.

Referring to the data processing pipeline of Higashi[27], cell filtering was performed on all datasets, excluding cells with fewer than 2,000 total reads or a genome span smaller than 500 kb. After preprocessing, Ramani2017 dataset consists of 620 cells from four human cell types, Lee2019 dataset consists of 4,238 cells from 14 human cell types, Tan2021A dataset consists of 1,692 cells from seven mouse cell types, Tan2021B dataset consists of 1,954 cells from 14 mouse cell types, and Wu2024 dataset consists of 431 cells from four human cell types.

The processed read pairs were binned into contact matrices at a specific genomic resolution. In this work, we only retained contact pairs within the same chromosome. The raw read data for all chromosomes of a cell were transformed into symmetric matrices of varying sizes, which served as the input for the model. For datasets with significant batch effects, such as Lee2019 dataset, we applied BandNorm suggested by Zheng et al.[28] to the data. In the dropout experiment, we randomly discarded raw read pairs with a certain probability, and then performed the processing steps mentioned above.

Downstream analysis

Evaluating low-dimensional cell embeddings. We evaluated the low-dimensional cell embeddings generated by Hi-Cformer on five publicly available scHi-C datasets with annotated cell types, all processed at 1-Mb resolution. To benchmark its performance, we compared Hi-Cformer against six established baseline methods: Higashi[27], scDEC-Hi-C[29], HiCRep/MDS[26], scHiCluster[23], PCA[34], and LDA[24]. Each baseline method was implemented following its respective publication and official codebase, using either recommended or default hyperparameter settings to ensure fairness. For Higashi, we used the version corresponding to the commits on October 11, 2021, from its GitHub repository. To ensure comparability, the dimensionality of the embeddings used for evaluation was set to 64 across all methods, except for scDEC-Hi-C. Due to the instability of its GAN-based

architecture, we were unable to obtain a converged model in its second-stage training under default settings. Therefore, for scDEC-Hi-C, we used the concatenated chromosome-wise embeddings generated from its first-stage autoencoder as the final cell representation, resulting in a higher-dimensional embedding of 1150 dimensions. For each method, we applied Leiden clustering to assess their ability to capture cell type heterogeneity. To align the number of clusters with the ground-truth cell type annotations, we employed a binary search strategy to select the clustering resolution. The clustering performance was quantified using NMI and ARI. The detailed formulas for computing these metrics are available in Supplementary Note A.1.

To assess the robustness of Hi-Cformer to data sparsity, we simulated dropout in the Ramani2017 dataset by applying random contact removal directly to the original paired-end read files. Each contact was independently retained with a fixed probability 1−p, where p is the dropout rate. The resulting contact pairs were then aggregated into single-cell contact frequency matrices at 1-Mb resolution, following the same preprocessing pipeline as in the main analysis. All model settings, training procedures and evaluation criteria remained consistent with those used in the original embedding experiment on the Ramani2017 dataset, ensuring that the robustness evaluation focused solely on the effect of data sparsity.

To evaluate the contribution of each component of Hi-Cformer, we conducted ablation experiments by removing the transformer module, preheating strategy, and MLM task. Three variants of the model were created by sequentially omitting these components, and the embeddings generated by each model were compared to those of the original Hi-Cformer model using the Ramani2017 dataset.

Data imputation and visualization. To evaluate the ability of Hi-Cformer to impute scHi-C contact maps, we compared it with two other imputation methods, Higashi[27] and scHiCluster[23]. In the section "Hi-Cformer provides reliable imputation of scHi-C contact maps", the model parameters were set in the same way as for the cell embedding evaluation. The reconstructed cell signals, denoted as $\boldsymbol{r}^{\text{cell}}$, generated by Hi-Cformer were considered as the imputed results. For similarity comparison, we computed the Pearson correlation coefficient (PCC) and cosine similarity between the imputed contact maps and the corresponding bulk Hi-C maps. We also calculated these two metrics between the imputed single-cell maps and the pseudo-bulk maps, which were obtained by averaging scHi-C contact maps from the same cell type. To avoid redundancy, the upper triangular matrix of each contact map was taken and flattened for the calculation. For the dropout experiments, we generated datasets with varying dropout ratios in the same way as for the cell embedding evaluation, using the raw scHi-C data as the ground truth. The imputation accuracy was measured using the PCC, cosine similarity, structural similarity index measure (SSIM), and peak signal-to-noise ratio (PSNR), with the respective formulas provided in Supplementary Note A.2. PCA was applied to reduce the dimensionality of the imputed maps to 64 components, followed by clustering to evaluate the effectiveness of imputation in revealing cell-type heterogeneity.

In the section "Hi-Cformer facilitates the identification of cell-type specific structures", we introduced a specialized 50-kb resolution region (chr9: 100,000,000-140,000,000) as input. Specifically, this region was passed through an additional fully connected layer to transform it into a 128-dimensional embedding. To distinguish this region from others, a unique size embedding was added to it. In the attention mechanism, the embedding of this special region was configured to

calculate the attention score with its corresponding chromosomal map embedding, while preventing it from being used by other embeddings for attention score computation. Finally, this specialized embedding passed through the transformer module and was processed by an added deconvolutional network to reconstruct the signal. The reconstructed signal, treated as the imputation result, was compared to the original input to compute the Mean Squared Error (MSE) loss, which was incorporated into the Loss_{block} for joint optimization. For the visualization of the Pearson correlation matrix, we applied the observed–expected normalization method introduced by Lieberman-Aiden et al.[5] before performing the calculation. The methods for calling TAD-like boundaries and A/B compartments followed the approach proposed in Higashi.

Supervised cell type annotation. Hi-Cformer was extended to support supervised cell type annotation by introducing a simple structural modification. Specifically, we added a single-layer fully connected neural network to the original Hi-Cformer framework, which takes the cell embedding $\boldsymbol{e}'^{\mathrm{cell}}$ as input and produces multiclass predictions using a softmax activation. The cross-entropy loss from this supervised task was incorporated into the original Hi-Cformer loss function.

We compared Hi-Cformer with three baseline methods: scHiClassifier[30], logistic regression (LR), and random forest (RF). For LR and RF, we used the four feature sets extracted from scHi-C contact maps as proposed by Zhou et al.[30] All methods were evaluated on five datasets using five-fold cross-validation. Classification performance was assessed using three metrics: Accuracy, Cohen's Kappa, and Macro-F1, with details of metric computation provided in Supplementary Note A.3. For cross-dataset evaluation between the Tan2021A and Tan2021B datasets, we performed label alignment to enable meaningful comparison. Specifically, in the evaluation phase, several neuronal subtypes in Tan2021A were merged into a single "Neuron" category (see Supplementary Note A.4 for the exact cell types), while the original labels were retained during training. This means that when a model trained on Tan2021A dataset is applied to Tan2021B dataset, the neuron subtype prediction (such as Neonatal Neuron 1) it gives will be treated as the "Neuron" class. When a model trained on the Tan2021B dataset is applied to the Tan2021A dataset, its prediction is considered correct if it classifies a neuronal subtype as the general "Neuron" category. To ensure fair comparison, cell types that were not present in the training set were excluded from metric calculation in the test set.

Data availability

We used the following publicly available datasets: scHi-C dataset of four human cell lines (GM12878, HAP1, HeLa and K562) from Ramani et al.[15] (GEO: GSE84920); sn-m3c-seq of human prefrontal cortex cells from Lee et al.[16] (GSE130711); scHi-C dataset (Dip-C) of mouse brain development was collected from Tan et al.[17] (GEO: GSE164203 and GEO: GSE146397); scHi-C dataset of four human cell lines (BJ, GM12878, K562 and eHAP) from Wu et al.[18] (GEO: GSE240128); bulk Hi-C of K562 and GM12878 from Rao et al.[14] (GEO: GSE63525); and ChIP-seq of H3K27ac in GM12878 and K562 from ENCODE[40] (GEO: GSE29611).

Code availability

Hi-Cformer is freely available on GitHub (https://github.com/Xiaoqing-Wu02/Hi-Cformer).

Acknowledgements

This work was supported by the National Key Research and Development Program of China (grant nos. 2023YFF1204802 and 2022YFF1202400) and the National Natural Science Foundation of China (grant nos. 62273194).

Author contributions

R.J. conceived the study and supervised the project. X.W. and X.C. designed, implemented and validated Hi-Cformer. X.W., X.C., and R.J. wrote the manuscript, with input from all the authors.

Competing interests

The authors declare no competing interests.

References

1. Tolhuis, B., Palstra, R.-J., Splinter, E., Grosveld, F. & De Laat, W. Looping and interaction between hypersensitive sites in the active β-globin locus. *Mol. Cell* **10**, 1453–1465 (2002).
2. Tang, Z. *et al.* CTCF-mediated human 3D genome architecture reveals chromatin topology for transcription. *Cell* **163**, 1611–1627 (2015).
3. Dekker, J. *et al.* The 4D nucleome project. *Nature* **549**, 219–226 (2017).
4. Marchal, C., Sima, J. & Gilbert, D. M. Control of DNA replication timing in the 3D genome. *Nat. Rev. Mol. Cell Biol.* **20**, 721–737 (2019).
5. Lieberman-Aiden, E. *et al.* Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* **326**, 289–293 (2009).
6. Lupiáñez, D. G. *et al.* Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012–1025 (2015).
7. Franke, M. *et al.* Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature* **538**, 265–269 (2016).
8. Bonev, B. & Cavalli, G. Organization and function of the 3D genome. *Nat. Rev. Genet.* **17**, 661–678 (2016).
9. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing Chromosome Conformation. *Science* **295**, 1306–1311 (2002).
10. Duan, Z. *et al.* A three-dimensional model of the yeast genome. *Nature* **465**, 363–367 (2010).
11. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
12. Sexton, T. *et al.* Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell* **148**, 458–472 (2012).
13. Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
14. Rao, S. S. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
15. Ramani, V. *et al.* Massively multiplex single-cell Hi-C. *Nat. Methods* **14**, 263–266 (2017).

16. Lee, D.-S. *et al.* Simultaneous profiling of 3D genome structure and DNA methylation in single human cells. *Nat. Methods* **16**, 999–1006 (2019).
17. Tan, L. *et al.* Changes in genome architecture and transcriptional dynamics progress independently of sensory experience during post-natal brain development. *Cell* **184**, 741–758 (2021).
18. Wu, H. *et al.* Simultaneous single-cell three-dimensional genome and gene expression profiling uncovers dynamic enhancer connectivity underlying olfactory receptor choice. *Nat. Methods* **21**, 974–982 (2024).
19. Nagano, T. *et al.* Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **502**, 59–64 (2013).
20. Schmitt, A. D., Hu, M. & Ren, B. Genome-wide mapping and analysis of chromosome architecture. *Nat. Rev. Mol. Cell Biol.* **17**, 743–755 (2016).
21. Zhang, Y. *et al.* Computational methods for analysing multiscale 3D genome organization. *Nat. Rev. Genet.* **25**, 123–141 (2024).
22. Dautle, M. A. & Chen, Y. Single-Cell Hi-C Technologies and Computational Data Analysis. *Adv. Sci.* **12**, 2412232 (2025).
23. Zhou, J. *et al.* Robust single-cell Hi-C clustering by convolution- and random-walk–based imputation. *Proc. Natl. Acad. Sci.* **116**, 14011–14018 (2019).
24. Kim, H.-J. *et al.* Capturing cell type-specific chromatin compartment patterns by applying topic modeling to single-cell Hi-C data. *PLoS Comput. Biol.* **16**, e1008173 (2020).
25. Zhang, R., Zhou, T. & Ma, J. Ultrafast and interpretable single-cell 3D genome analysis with Fast-Higashi. *Cell Syst.* **13**, 798–807 (2022).
26. Li, X., Feng, F., Pu, H., Leung, W. Y. & Liu, J. scHiCTools: A computational toolbox for analyzing single-cell Hi-C data. *PLoS Comput. Biol.* **17**, e1008978 (2021).
27. Zhang, R., Zhou, T. & Ma, J. Multiscale and integrative single-cell Hi-C analysis with Higashi. *Nat. Biotechnol.* **40**, 254–261 (2022).
28. Zheng, Y., Shen, S. & Keleş, S. Normalization and de-noising of single-cell Hi-C data with BandNorm and scVI-3D. *Genome Biol.* **23**, 222 (2022).
29. Liu, Q. *et al.* Deep generative modeling and clustering of single cell Hi-C data. *Brief. Bioinform.* **24**, bbac494 (2023).
30. Zhou, X. & Wu, H. scHiClassifier: a deep learning framework for cell type prediction by fusing multiple feature sets from single-cell Hi-C data. *Brief. Bioinform.* **26**, bbaf009 (2025).
31. Vaswani, A. *et al.* Attention is all you need. *Adv. Neural Inf. Process. Syst.* **30**, (2017).
32. Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. in *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)* 4171–4186 (2019).
33. Yang, T. *et al.* HiCRep: assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient. *Genome Res.* **27**, 1939–1949 (2017).
34. Hotelling, H. Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* **24**, 417 (1933).

35. Blei, D. M., Ng, A. Y. & Jordan, M. I. Latent dirichlet allocation. *J. Mach. Learn. Res.* **3**, 993–1022 (2003).
36. Traag, V. A., Waltman, L. & Van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 1–12 (2019).
37. Nagano, T. *et al.* Cell-cycle dynamics of chromosomal organization at single-cell resolution. *Nature* **547**, 61–67 (2017).
38. Ren, R. Mechanisms of BCR–ABL in the pathogenesis of chronic myelogenous leukaemia. *Nat. Rev. Cancer* **5**, 172–183 (2005).
39. Bintu, B. *et al.* Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science* **362**, eaau1783 (2018).
40. Consortium, E. P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57 (2012).
41. Wu, H. *et al.* Simultaneous single-cell three-dimensional genome and gene expression profiling uncovers dynamic enhancer connectivity underlying olfactory receptor choice. *Nat. Methods* **21**, 974–982 (2024).

Supplementary Information for

Hi-Cformer enables multi-scale chromatin contact map modeling for single-cell Hi-C data analysis

Xiaoqing Wu[1,†], Xiaoyang Chen[1,†], Rui Jiang[1,*]

[1] Ministry of Education Key Laboratory of Bioinformatics, Bioinformatics Division at the Beijing National Research Center for Information Science and Technology, Center for Synthetic and Systems Biology, Department of Automation, Tsinghua University, Beijing 100084, China

† These authors contributed equally to this work.

* To whom correspondence should be addressed. Email: ruijiang@tsinghua.edu.cn

Contents

A Supplementary Notes 3

A.1 Evaluation metrics for cell embedding clustering 3

A.2 Evaluation metrics for imputation accuracy 4

A.3 Evaluation metrics for classification performance 5

A.4 Label alignment for cross-dataset classification evaluation 6

B Supplementary Figures 7

Supplementary Figure S1 7

Supplementary Figure S2 8

Supplementary Figure S3 9

A Supplementary Notes

A.1 Evaluation metrics for cell embedding clustering

To assess the consistency between predicted clusters and ground-truth labels, we used two widely adopted clustering evaluation metrics: normalized mutual information (NMI) and adjusted rand index (ARI).

Let $C = \{C_1, C_2, \dots, C_K\}$ be the set of inferred cell type labels by Leiden algorithm, and $G = \{G_1, G_2, \dots, G_J\}$ the set of ground-truth cell type labels. Denote n as the total number of samples.

NMI measures the mutual dependence between C and G, and is defined as:

$$\mathrm{NMI}(C, G) = \frac{2 \cdot I(C; G)}{H(C) + H(G)}$$

where $I(C; G)$ is the mutual information between the inferred and the ground-truth labels, and $H(C)$ and $H(C)$ are the entropies of the inferred and true label distributions, respectively.

$$I(C; G) = \sum_{i=1}^{K} \sum_{j=1}^{J} \frac{|C_i \cap G_j|}{n} \log\left(\frac{n \cdot |C_i \cap G_j|}{|C_i| \cdot |G_j|}\right)$$

$$H(C) = -\sum_{i=1}^{K} \frac{|C_i|}{n} \log\left(\frac{|C_i|}{n}\right)$$

$$H(G) = -\sum_{j=1}^{J} \frac{|G_j|}{n} \log\left(\frac{|G_j|}{n}\right)$$

ARI measures the agreement between cell type labels inferred by the Leiden algorithm and the ground truth labels by considering all pairs of samples and counting pairs that are assigned consistently in both C and G. It adjusts for chance grouping:

$$\mathrm{ARI}(C, G) = \frac{\sum_{ij}\binom{n_{ij}}{2} - \left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2} / \binom{n}{2}\right]}{\frac{1}{2}\left[\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2}\right] - \left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2} / \binom{n}{2}\right]}$$

where $n_{ij} = |C_i \cap G_j|$, $a_i = |C_i|$, and $b_j = |G_j|$.

Both NMI and ARI range from 0 to 1, with higher values indicating better agreement between inferred and true labels.

A.2 Evaluation metrics for imputation accuracy

To evaluate the imputation accuracy of single-cell Hi-C contact maps, we used the following quantitative metrics: area under the Precision-Recall curve (AUPR), Pearson correlation coefficient (PCC), cosine similarity, structural similarity index measure (SSIM), and peak signal-to-noise ratio (PSNR). Let $X = (x_1, x_2, \dots, x_n)$ and $Y = (y_1, y_2, \dots, y_n)$ denote the vectorized upper triangles of two contact matrices to be compared (e.g., imputed vs. ground truth).

The PCC between X and Y is defined as:

$$\mathrm{PCC}(X,Y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

where $\bar{x}$ and $\bar{y}$ denote the sample means of X and Y, respectively.

Cosine similarity measures the cosine of the angle between two vectors:

$$\mathrm{Cosine}(X,Y) = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2}\sqrt{\sum_{i=1}^{n} y_i^2}}$$

Given two contact matrices reshaped into square matrices X' and Y', SSIM is computed as:

$$\mathrm{SSIM}(X',Y') = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

where μ_x and μ_y are the means, σ_x^2 and σ_y^2 are the variances, and σ_{xy} is the covariance. $C_1 = 0.0001$ and $C_2 = 0.0009$ are small constants for numerical stability.

Assuming matrices X', Y' of size $m \times m$, the PSNR is defined as:

$$\mathrm{PSNR}(X',Y') = 10\log_{10}\left(\frac{\mathrm{MAX}^2}{\mathrm{MSE}(X',Y')}\right)$$

where MAX is the maximum possible pixel value (1 after normalization), and MSE is the mean squared error:

$$\mathrm{MSE}(X',Y') = \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}\left(X'_{ij} - Y'_{ij}\right)^2$$

Prior to computing SSIM and PSNR, all matrices were min-max normalized such that their entries were scaled to the range $[0,1]$. This normalization ensures consistency in dynamic range across different contact maps and avoids scale-induced distortions in similarity measurements.

A.3 Evaluation metrics for classification performance

To assess the classification performance of the methods, we employed three evaluation metrics: accuracy, Cohen's kappa, and macro-F1 score. Let $y_{\text{true}} = (y_1, y_2, \dots, y_n)$ represent the true cell type labels and $y_{\text{pred}} = (y'_1, y'_2, \dots, y'_n)$ represent the predicted cell type labels, where n is the total number of cells.

Accuracy is defined as the proportion of correctly annotated instances to the total number of instances:

$$\text{Accuracy} = \frac{1}{n}\sum_{i=1}^{n} 1(y_i = y'_i)$$

where $1(\cdot)$ is the indicator function, which equals 1 if the condition holds true and 0 otherwise.

Cohen's kappa measures the agreement between predicted and true cell type labels while accounting for the possibility of agreement occurring by chance. It is computed as:

$$\kappa = \frac{P_{\text{observed}} - P_{\text{expected}}}{1 - P_{\text{expected}}}$$

where P_{observed} is the observed agreement, and P_{expected} is the expected agreement by chance. Both are computed using the confusion matrix:

$$P_{\text{observed}} = \frac{1}{n}\sum_{i=1}^{n} 1(y_i = y'_i)$$

$$P_{\text{expected}} = \sum_{k=1}^{K} \left(\frac{C_{kk}}{n}\right)^2$$

where C_{kk} represents the count of correctly predicted instances for cell type k, and K is the total number of cell types.

Macro-F1 score is defined as the unweighted mean of per-cell-type F1 scores:

$$\text{F1}_{\text{macro}} = \frac{1}{K}\sum_{k=1}^{K} \frac{2 \cdot \text{Precision}_{\text{k}} \cdot \text{Recall}_{\text{k}}}{\text{Precision}_{\text{k}} + \text{Recall}_{\text{k}}}$$

where K is the total number of classes, $\text{Precision}_{\text{k}} = \frac{TP_k}{TP_k + FP_k}$ is the precision for cell type k, and $\text{Recall}_{\text{k}} = \frac{TP_k}{TP_k + FN_k}$ is the recall for cell type k. TP_k, FP_k, and FN_k denote the number of true positives, false positives, and false negatives for cell type k, respectively. Macro-F1 treats each cell type equally, regardless of cell type imbalance.

A.4 Label alignment for cross-dataset classification evaluation

To enable consistent and meaningful evaluation across datasets with differing label granularities, we performed label alignment in the intra-dataset annotation experiments between the Tan2021A dataset and Tan2021B dataset. Specifically, several neuronal subtypes annotated in the Tan2021A dataset were merged into a single generalized "neuron" cell type during the evaluation phase. This was done to match the coarser cell type labels used in the Tan2021B dataset.

The relabeling only applied at test time for evaluation purposes. During training, the original fine-grained labels from Tan2021A were retained to preserve subtype-specific information.

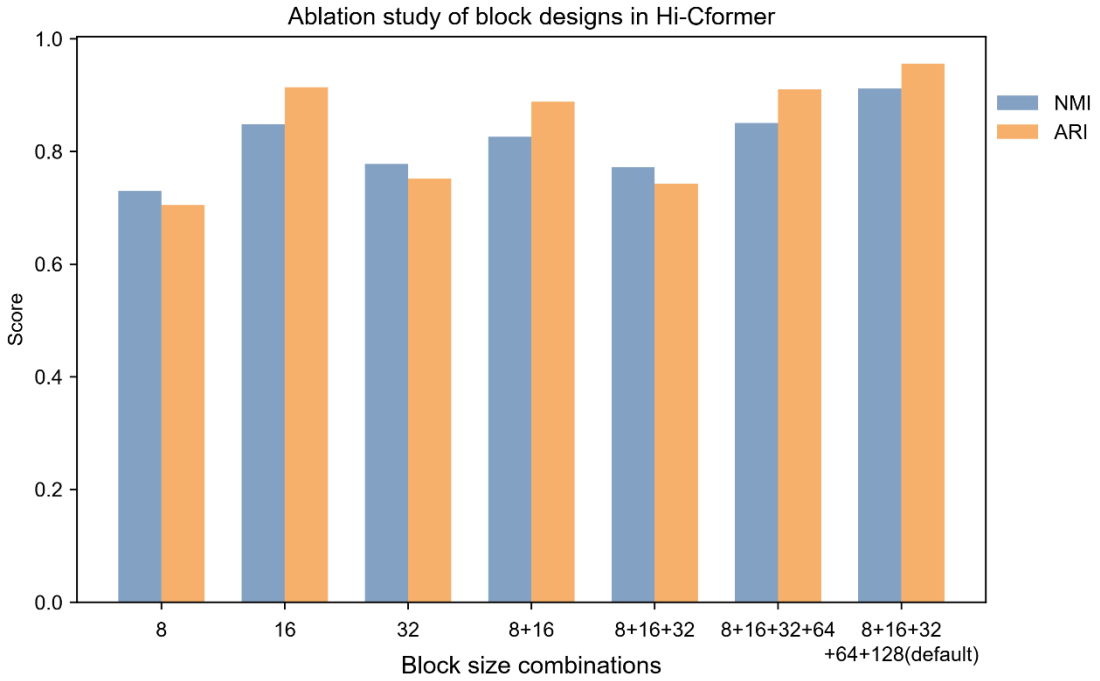
The neuron subtypes merged into the unified "neuron" class include:

- Cortical L2-5 Pyramidal Cell
- Cortical L6 Pyramidal Cell
- Hippocampal Granule Cell
- Hippocampal Pyramidal Cell
- Medium Spiny Neuron
- Neonatal Neuron 1
- Neonatal Neuron 2

When a model trained on Tan2021A was tested on Tan2021B, any prediction corresponding to one of these subtypes was counted as a correct annotation for the "neuron" cell type, if the true label in Tan2021B was also "neuron". This alignment ensured that differences in annotation resolution between datasets did not artificially penalize performance.

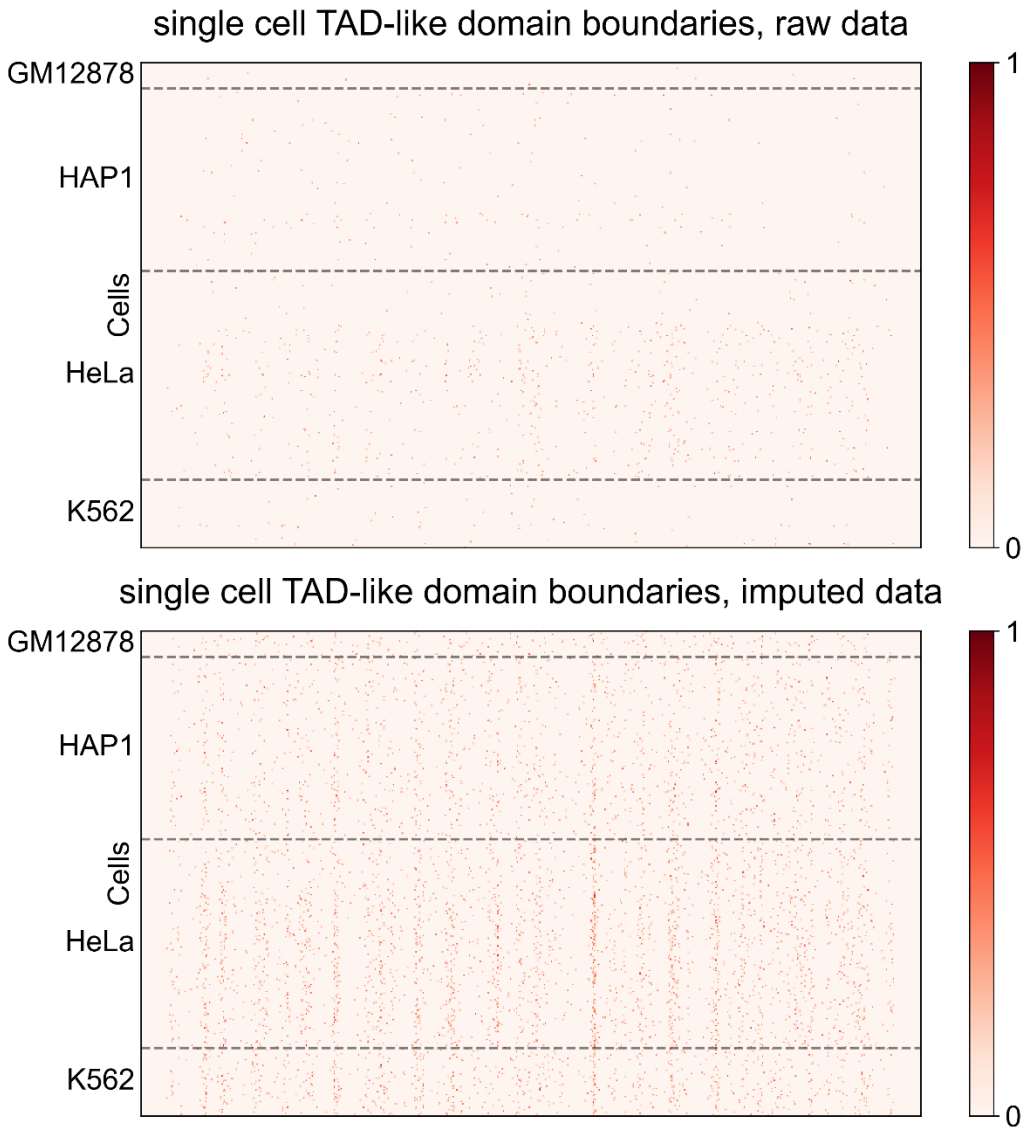
B Supplementary Figures

Supplementary Figure S1



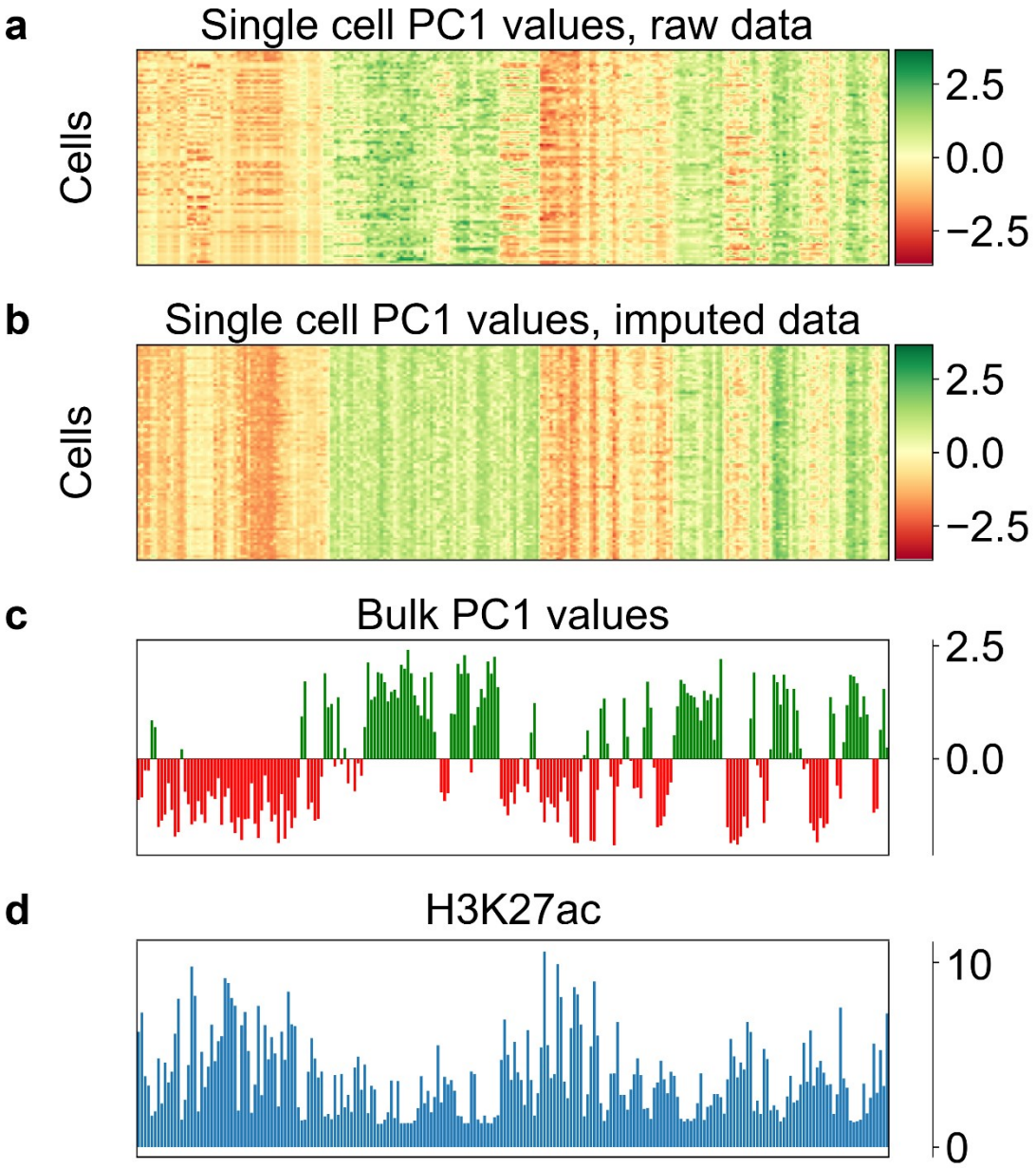
Supplementary Figure S1 | Ablation study of block designs in Hi-Cformer. Clustering performance of Hi-Cformer under different block size combinations on the Ramani2017 dataset, evaluated using NMI and ARI. The x-axis denotes the block sizes used in the multi-scale encoder, ranging from single-scale settings (e.g., 8, 16, or 32) to progressively deeper multi-scale combinations (e.g., 8+16, 8+16+32, etc.). The final configuration (8+16+32+64+128), marked as default, includes the full set of scales.

Supplementary Figure S2



Supplementary Figure S2 | Detection of TAD-like domain boundaries from raw and imputed scHi-C data. Binary boundary maps derived from raw and imputed single-cell Hi-C contact matrices for four cell types, visualized along chr9:100,000,000–140,000,000 at 50-kb resolution. Each row corresponds to a single cell, and each column represents a genomic bin; red marks indicate predicted TAD-like boundaries. Color bars denote binary boundary presence (1) or absence (0). Cell type groupings are separated by dashed lines.

Supplementary Figure S3



Supplementary Figure S3 | Detection of A/B compartments from raw and imputed single-cell Hi-C data. a, PC1 values computed from raw single-cell Hi-C contact matrices of K562 cells, representing A/B compartment scores of chr1:0–249,250,621 at 1-Mb resolution. Each row corresponds to one cell, and each column to a genomic bin. **b**, PC1 values derived from imputed single-cell contact matrices, showing enhanced compartmentalization patterns across cells. **c**, Bulk-level PC1 values computed from aggregated Hi-C data over the same genomic region, serving as a reference. **d**, ChIP-seq signal intensities for H3K27ac of the same region. All tracks are shown at matched genomic resolution to facilitate direct comparison across single-cell, bulk, and epigenomic profiles.