BMB
Reports

BMB Rep. 2025; 58(12): 485-493
www.bmbreports.org

Invited Mini Review

Recent advances in single-cell bioinformatics for inferring higher-order chromatin contact maps

Seung Kyun Noh[#], *Minhyeok Lee*[#] *& Hyobin Jeong*[*]
Department of Systems Biology, College of Life Science and Biotechnology, Yonsei University, Seoul 03722, Korea

DNA, a large molecule located in the nucleus, carries essential genetic information, including gene loci and cis-regulatory elements. Despite its extensive length, DNA is compactly stored within the limited space of the nucleus due to its hierarchical three-dimensional (3D) organization. In this structure, DNA is organized into territories known as topologically associated domains (TADs). Within each TAD, numerous chromatin loops link promoters and enhancers across the genome. These loops and the interactions between promoters and enhancers are dynamically regulated, thereby controlling gene transcription activities. With the rapid advancements in single-cell genomics technologies, TAD boundaries and chromatin loops can now be observed at the level of individual cells, allowing researchers to explore cellular heterogeneity in tissues. This review will summarize the state-of-the-art bioinformatics methods recently developed to analyze single-cell Hi-C and epigenomics datasets, which infer higher-order chromatin interactions within the 3D genome. Additionally, we will discuss the biological applications of these tools and future directions for comprehensively investigating epigenomic heterogeneity across different species, developmental stages, and disease states. [BMB Reports 2025; 58(12): 485-493]

INTRODUCTION

Importance of exploring higher-order chromatin contact maps in single-cells

What is the definition of a higher-order chromatin contact map?: The genomes of diverse species are often extremely long but are efficiently packed into the small nucleus of a cell due to the three-dimensional (3D) organization of the genome. Chromatin architecture facilitates this effective 3D organization, which is specialized for precise regulation of transcriptional activities and epigenetic controls in cells. A higher-order chromatin contact map captures and visualizes concurrent interactions between multiple genomic regions by definition. A higher-order chromatin contact map provides a detailed representation of the 3D organization and spatial interactions of chromatin within a cell's nucleus (1). These maps delineate chromatin folding patterns and the interactions between distant DNA regions, including loops, topologically associating domains (TADs), compartmentalization, and chromosome territories.

Why do we create a higher-order chromatin contact map?: Studying higher-order chromatin contact maps enhances our understanding of the communication between genes and regulatory elements across long distances within the genome. By generating a higher-order chromatin contact map, we can investigate the organization of chromosomal territories, TADs, and A/B compartments. A compartments contain active chromatin, whereas B compartments harbor inactive chromatin (2). Moreover, A/B compartments can be subdivided into five primary subcompartments: A1, A2, B1, B2, and B3, each with distinct genomic and epigenomic characteristics, such as varying levels of gene expression, active and repressive histone marks, DNA replication timing, and specific subnuclear structures (3).

Additionally, it can be employed to deduce promoter-enhancer looping and regulatory circuits of genes. The transcription of genes is tightly controlled by cis-regulatory elements, including promoters and enhancers. Typically, promoters are defined as regulatory elements that are proximal to the transcription start site (TSS) of a gene, activating transcription when transcription factors and RNA polymerase bind to their DNA sequences. Conversely, enhancers are distal regulatory elements situated relatively far from the TSS of genes (distance to TSS $>$ 1 kb). Cohesin and CTCF facilitate interactions between promoters and enhancers to precisely regulate the expression of target genes. Multiple enhancers may control the transcription of specific genes, and the selection of particular enhancers and promoters for each gene depends on the celltype and context. In some cases, the transcription of genes is regulated by long-range interactions that span distances greater than 50 kb, facilitated by 3D genome interactions. Constructing higher-order chromatin contact maps enables the systematic identification of interactions between promoters

*Corresponding author. Tel: +82-2-2123-2659; Fax: +82-2-312-5657; E-mail: hyobinjeong@yonsei.ac.kr
[#]These authors contributed equally to this work.

https://doi.org/10.5483/BMBRep.2025-0015

Received 3 February 2025, Revised 4 March 2025, Accepted 28 March 2025, Published online 14 May 2025

Keywords: Bioinformatics, Higher-order contact map, scATAC-seq, scHi-C, Single-cell

ISSN: 1976-670X (electronic edition)

and enhancers genome-wide under specific biological contexts.

Additionally, higher-order chromatin contact maps are instrumental in detecting the presence and locations of structural variants (SVs). SVs are characterized as genetic variants larger than 50 base pairs, encompassing copy number alterations (deletions, duplications), inversions, and translocations (4). Traditionally, SVs have been detected using methods such as whole genome sequencing (WGS) or whole exome sequencing (WES), which identify the presence of SVs through abnormal read depths, and discordant reads indicative of breakpoints (5). However, the detection of copy-neutral events like inversions and balanced translocations remains a challenge, as it relies on the identification of breakpoint-spanning reads. Chromatin contact mapping techniques, such as mate-pair sequencing (6) and Hi-C contact-map based methods (7), can complement these techniques. The underlying principle of these methods is that the presence of an SV induces abnormal interactions between genomic positions that are usually distant. Such interactions can be readily identified by analyzing higher-order chromatin contact maps and comparing them with maps generated from normal cells of the same celltype. Thus, higher-order chromatin contact maps serve as valuable tools not only for exploring epigenomic landscapes but also for uncovering hidden structural rearrangements in the genome.

How can we generate higher-order chromatin contact maps in tissues and cell lines?: Traditionally, higher-order chromatin contact maps are constructed using experimental techniques that capture chromatin interactions within the three-dimensional nuclear space, followed by computational analyses to generate the contact maps. Various chromatin contact mapping techniques, including Hi-C, 3C, ChIA-PET, and capture Hi-C, measure the contact frequency between two genomic positions (8). More recently, innovative methods based on machine learning (ML) and artificial intelligence (AI) have been employed to infer chromatin contact maps from multiple bulk epigenomic datasets across different celltypes (Fig. 1) (9, 10).

Why is a single-cell higher-order chromatin map needed? How can we generate them?: As outlined in the preceding section, researchers have produced higher-order chromatin maps across various cell lines and tissues. These investigations have demonstrated that chromatin contact maps and higher-order chromatin interactions are markedly cell-type specific, underscoring the importance of examining higher-order contact maps at the single-cell level. Recent strides in genomics have facilitated the analysis of different omics layers at the single-cell level, as opposed to bulk analysis, enabling researchers to delineate cellular heterogeneity with greater resolution. This approach commenced with the profiling of single-cell transcriptomes and has since expanded to include multiple layers, such as single-cell genomes, single-cell chromatin accessibility, and single-cell chromatin contact maps (11, 12).

In this era of single-cell omics, we are able to construct higher-order chromatin maps using single-cell Hi-C and bioinformatics analyses, which directly quantify the interaction strengths between pairs of genomic regions on a genome-wide scale. Additionally, higher-order chromatin maps can be predicted via other epigenomic modalities such as single-cell ATAC-seq, and Bisulfite-seq based methods, by deducing chromatin interactions between two genomic regions based on the correlation of chromatin accessibility across multiple single cells (Fig. 1). With these technologies, we now possess the capability to explore cell-type-specific chromatin contact maps. Furthermore, these maps have the potential to enhance our understanding of the initiation and progression of diseases, including cancer (13). Although this approach has not yet been extensively explored, it could theoretically serve to delineate subclone-specific chromatin contact maps in cancer studies, where intra-tumor heterogeneity significantly contributes to cancer progression and drug resistance.

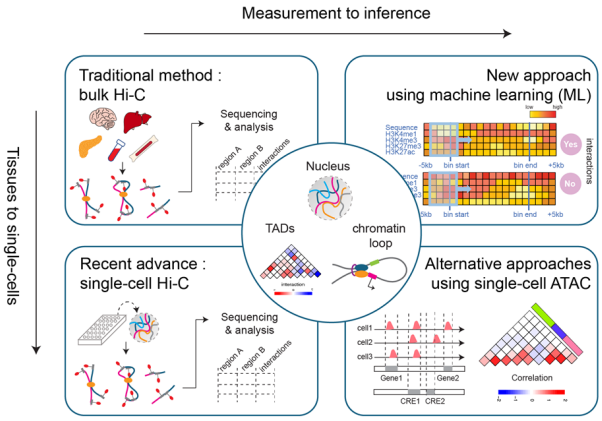

Fig. 1. Technologies to explore higher-order chromatin structure. Currently, available approaches for investigating higher-order chromatin structures are categorized into four primary groups. Methods on the left and right sides are distinguished based on whether they directly measure (left) or infer (right) higher-order chromatin structures. Methods in the upper and lower panels are differentiated by the scale of investigation, either at the tissue-level (upper) or at the single-cell-level (lower).

In this review, we will summarize the current state of single- cell bioinformatics methods for deriving higher-order chromatin contact maps utilizing single-cell Hi-C data sets and other types of single-cell epigenomic data sets, from the initial data acquisition to the final analytical steps that identify and visualize chromatin contact maps. In doing so, we aim to offer comprehensive comparisons of the currently available methods for analyzing single-cell higher-order chromatin maps and to propose potential avenues for future development aimed at enhancing our understanding of cellular heterogeneity through the analysis of higher-order chromatin maps in diverse biological systems.

RESULTS

Single-cell bioinformatics methods to infer higher-order chromatin contact maps

Single-cell Hi-C based bioinformatics methods: Upon obtaining data from single-cell Hi-C technology, the initial step involves preprocessing the raw data. This includes aligning the reads with the reference genome and generating a contact matrix prior to conducting downstream analysis for particular objectives (Fig. 2). Despite the diversity of experimental methods utilized in scHi-C library preparation, a standardized unified workflow for data processing is lacking. Nevertheless, preprocessing of scHi-C data typically encompasses seven fundamental analysis steps: 1) Demultiplexing into singlecells (optional, depending on the experimental protocol employed during library preparation), 2) Conducting quality checks on FASTQ files using FastQC (14), 3) Mapping reads to the reference genome using bwa-mem (15), 4) Generating the contact file, 5) Merging and sorting the contact files, 6) Filtering the contact files, and 7) Constructing the contact matrix.

Numerous bioinformatics software tools have been specifically developed for preprocessing single-cell Hi-C datasets.

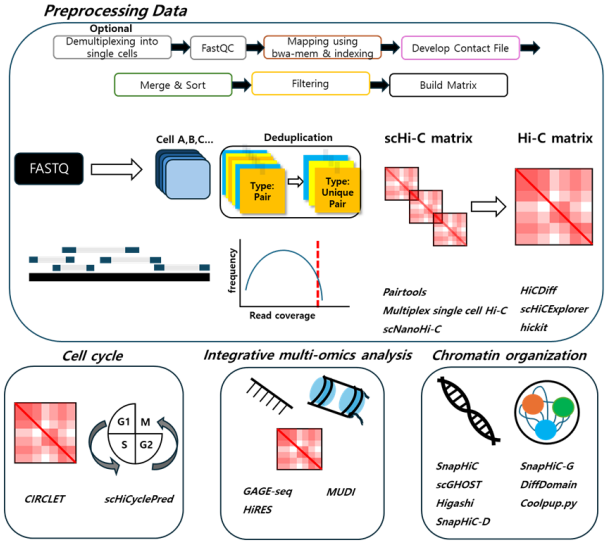

Fig. 2. Single-cell Hi-C based computational methods to infer chromatin contact map. Bioinformatic analysis of scHi-C packages begins with preprocessing steps involving bioinformatics packages necessary for demultiplexing, quality control, mapping, and enhanced visualization. Firstly, 'The cell cycle category' includes packages that identify genomic regions associated with cancer marker genes at the resolution of a single cell and also predict cell cycle phases using scHi-C data to characterize the dynamics of chromosome structure. Secondly, "The integrative multi-omics analysis" comprises packages that integrate scHi-C data with scRNAseq, facilitating single-cell multi-omics analysis. Thirdly, 'The chromatin organization category' includes packages designed to accurately reconstruct 3D chromatin contact maps using the scHi-C interaction score.

For example, Multiplex single-cell Hi-C (16) aids in the demultiplexing of single-cell Hi-C data derived through combinatorial cellular indexing. scHiCExplorer (17) executes general preprocessing steps for scHi-C data, encompassing demultiplexing, mapping to the reference genome, constructing interaction matrices, merging matrices, conducting quality control, performing normalization, and applying correction. Additionally, for single-cell Hi-C data, it offers specialized tools such as unsupervised clustering of single cells, and computing the A/B compartments of individual cells and clusters. HiCDiff (18) attenuates noise in scHi-C data through a diffusion model that can be trained in both unsupervised and supervised learning modes by incrementally adding Gaussian noise to the Hi-C data during training. Pairtools (19) manages the overall preprocessing of both bulk Hi-C and scHi-C data. scNanoHi-C (20) leverages the Snakemake workflow management system to streamline the preprocessing of single-cell long-read Hi-C sequencing data. Moreover, hickit (21) processes diploid single-cell Hi-C data. It optimizes and simplifies the Dip-C pipeline, facilitating rapid access to its functions. Hickit is capable of imputing missing phases and inferring 3D structures for each haplotype in the human diploid genome, thereby enabling the analysis of haplotype-specific 3D genome structures.

Following the preprocessing step, we can undertake three major downstream analyses with scHi-C data: exploring cell cycle phases, integrating multi-omic readouts, and reconstructing chromatin organization (Fig. 2). To analyze cell cycle phases with scHi-C data, tools such as CIRCLET (22), and scHiCyclePred (23) have been developed. CIRCLET (22) employs KNN graphs to reconstruct the cell cycle trajectory (G1, ES, MS, G2) from scHi-C contact matrices without the need to specify a starting cell. Conversely, scHiCyclePred (23) predicts cell cycle phases accurately by incorporating three sets of features extracted from scHi-C data and applying a Convolutional Neural Network (CNN) model based on multi-feature fusion.

Secondly, to integrate scHi-C with multi-omic readouts, GAGE-seq (24), MUDI (25), and HiRES (26) have been developed. These tools are specifically designed to integrate scHi-C datasets with other omic modalities during single-cell Hi-C profiling in multi-omic contexts. In this category, GAGE-seq (24) and HiRES (26) integrate multimodal datasets derived from scHi-C and scRNA co-profiling approaches, while MUDI (25) integrates scHi-C and scRNA-seq data independently profiled from different samples, by identifying 3D-regulated, biological-context-dependent cell subpopulations, known as topologically integrated subpopulations (TISPs).

Thirdly, for reconstructing chromatin organization and constructing higher-order chromatin contact maps, SnapHiC (27), SnapHiC-D (28), SnapHiC-G (29), scGHOST (30), Higashi (31), DiffDomain (32), and Collpup.py (33) are suitable (Fig. 2, Supplementary Table 1). As the primary focus of this review is on constructing higher-order chromatin contact maps, this discussion will particularly highlight the last category of bioinformatics tools designed to explore chromatin organi-

zation as a downstream analysis.

To explore chromatin organization using scHi-C, the primary challenge is the extreme sparsity of the data, necessitating the analysis of a large number of cells simultaneously. SnapHiC addresses this challenge by estimating intrachromosomal contact probabilities between pairs of 10 kb bins in each cell using the random walk with restart algorithm (27). In this method, individual cells are treated as independent datasets rather than being aggregated into pseudobulk, enhancing the statistical power of loop detection even when the number of cells is limited. Additionally, it quantifies the number of loops per cluster by chromosome.

Higashi (31) is an additional bioinformatics package designed to reconstruct 3D chromatin organization in individual single cells through hypergraph representation learning. This method exploits latent correlations among single cells to enhance the overall accuracy of contact map imputation. Higashi facilitates the generation of UMAP visualizations using embedding values, which assist in confirming distinct patterns of cell types and conditions. Consequently, cell-to-cell variability in A/B compartment scores and TAD-like domain boundaries can be effectively explored.

The 3D genome displays critical subcompartment patterns including A1, A2, B1, B2, and B3, which the binary A/B compartment definitions fail to capture. scGHOST (30), a single-cell subcompartment annotation method, employs graph embedding with constrained random walk sampling. This tool requires two inputs: the imputed scHi-C contact maps of a cell and scHi-C embeddings generated via Higashi. It subsequently identifies the k-nearest neighbors for each cell based on single-cell embeddings. Following this, it calculates genomic locus embeddings tailored for downstream clustering of single-cell subcompartments. Consequently, scGHOST reveals cell-to-cell variability in single-cell subcompartments and enables cell-type-specific genome structure analysis.

Not only can single-cell chromatin organization be reconstructed, but differential chromatin contacts (DCCs) between two scHi-C datasets can also be discerned using tools such as SnapHiC-D (28). SnapHiC-D utilizes the random walk with restart (RWR) algorithm, similar to the method described for SnapHi-C, to address the challenges of sparsity and low sequencing depth in single-cell data. As a result, the differential chromatin contacts identified by SnapHiC-D can elucidate cell-type-specific gene expression and epigenetic features at single-cell resolution. Additionally, SnapHiC-G can be applied to scHi-C data to identify long-range enhancer-promoter interactions using a global background model (29). SnapHiC-G also identifies potential target genes in GWAS variants (non-coding genome-wide association studies), thereby serving as a powerful tool for characterizing cell-type-specific enhancer-promoter interactions.

Diffdomain is a tool designed for identifying reorganized TADs using single cell Hi-C contact matrices from two distinct biological conditions. It classifies TAD reorganization into one of six subtypes: strength-change, merge, loss, split, zoom, and complex, which facilitate the characterization of features within reorganized chromatin domains.

coolpup.py is a versatile tool for conducting pile-up analyses using .cool files in Hi-C format. This tool quantifies the intensity of loops and TAD boundaries in specific regions through pile-up analysis. Furthermore, it facilitates the observation of interaction patterns between specific genomic regions, thereby aiding in the analysis of chromatin organization.

Single-cell ATAC-seq based bioinformatics method: Single-cell Hi-C, a pivotal single-cell epigenomics technology, directly measures higher-order chromatin interactions at the individual cell level. In addition to single-cell Hi-C, a range of other single-cell epigenomic technologies are now available, including scATAC-seq and scMNase-seq, which assess genome-wide chromatin accessibility, and scMethyl-seq, which evaluates genome-wide DNA methylation profiles. scATAC-seq, frequently utilized due to its accessibility via the commercial 10X genomics platform, is especially prominent for assessing cellular heterogeneity in chromatin accessibility, clustering and annotating celltypes, and identifying cell-type specific motif activities (34). scATAC-seq also offers potential for inferring long-range interactions between cis-regulatory elements on a genome-wide scale, providing critical data for reconstructing higher-order chromatin contact maps or loops. This section focuses on the scATAC-seq bioinformatics workflow for inferring 3D genome structure as an alternative to scHi-C based approaches.

Upon obtaining scATAC-seq data in FASTQ format, we align the reads to the reference genome using CellRanger-atac (35) to generate fragment files and BAM files. Subsequent analyses include preprocessing, dimensionality reduction, clustering, peak calling, differential analysis, motif analysis, and data integration through computational pipelines such as CellSpace (36), snapATAC2 (37), Signac (38), ArchR (39), and scATAC-pro (40) (Fig. 3). Each package provides a unique set of functionalities and offers an interface based on either the R or Python programming languages (Fig. 3).

After the main analysis, higher-order chromatin contact maps can be directly predicted using methods such as ChromaFold (41), Compartmap (42), or the correlation-based methodologies proposed by Buenrostro et al (43) (Supplementary Table 2, Fig. 3). In the approach described by Buenrostro et al (43), chromatin accessibility of non-overlapping 500 bp peaks is determined genome-wide. Subsequently, genomic bins comprising proximal accessibility peaks (designated as peak bins) are generated (N = 25 peaks representing a median size of 135 kb, with a step size of 10 peaks), followed by the calculation of a bin-level normalized deviation score relative to a background peak set. Thereafter, the correlation coefficient of each accessible peak bin with all other accessible peak bins across all cells is computed as a proxy for interaction frequency, analogous to measurements obtained by chromatin conformation capture assays such as Hi-C, thereby allowing inference of higher-order chromatin contact maps.

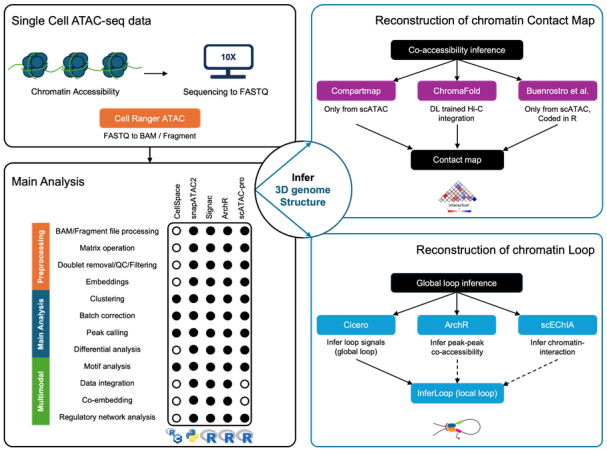

Fig. 3. Single-cell chromatin accessibility based computational methods to infer 3D chromatin structure. Single-cell ATAC-seq reads stored in input FASTQ files are aligned to the reference genome using CellRanger-ATAC, which generates fragment files and BAM files. Basic analyses, including pre-processing, dimensionality reduction, clustering, peak calling, and motif analysis, are performed using computational pipelines such as CellSpace, snapATAC2, Signac, ArchR, and scATAC-pro. Following these basic analysis steps, higher-order chromatin contact maps can be directly reconstructed through co-accessibility inference with methods like Compartmap (which utilizes scATAC-seq data exclusively), ChromaFold (employing a previously trained deep-learning model with Hi-C), and the method proposed by Buenrostro et al. Alternatively, cell-type specific local loops can be derived from global loop inference using methods such as Cicero. Other bioinformatics tools like ArchR and scEChIA can be employed to infer global loops. After inferring global loops, cell-type specific local loops can be deduced using InferLoop.

Alternatively, ChromaFold (41) employs a supervised deep-learning model to predict 3D contact maps from scATAC-seq data and CTCF motif tracks as input features. It offers two different pipeline modes: [ChromaFold + CTCF motif] and [ChromaFold + CTCF ChIP]. The former mode, which does not necessitate additional ChIP-seq experiments, provides broader applicability. To develop this prediction model, ChromaFold was trained using paired scATAC-seq and Hi-C data from a variety of training cell types.

Another technique, Compartmap, aims to estimate A/B compartments through eigenvector analysis of the genome contact matrix, normalized via the observed-expected method (42). Specifically, when applied to scATAC-seq data, it utilizes a James-Stein estimator (JSE) aimed at a global or targeted mean, leveraging either chromosome-level or genome-wide information from scATAC-seq. To mitigate the sparsity inherent in single-cell data, it employs a bootstrap procedure to quantify the uncertainty associated with the state and boundaries of inferred compartments. In its final step, it visualizes the inferred higher-order interacting domains using a Random Matrix Theory approach, producing plaid-like patterns reminiscent of those seen in Hi-C data.

Following the main analysis, there exists an alternative approach to predict cell-type-specific local loops through a two-step process: 1) inferring global loops, and 2) predicting local loops based on these global configurations. Initially, to predict cell-type-specific local loops, it is necessary to infer global loops using Cicero (44). Alongside Cicero, other bioinformatics tools capable of inferring chromatin interaction loops, such as ArchR (45), and scEChIA, are also applicable at this stage. Once global loops are predicted, local loops can be inferred using methodologies like InferLoop (46) (Supplementary Table 2). The input for InferLoop consists of two main components: firstly, a matrix of peak signals (with rows representing peaks and columns representing cell bins), and secondly, a list of loops previously predicted by pipelines such as Cicero (44). The InferLoop workflow initiates with signal enhancement by aggregating nearby cells into bins, wherein it employs accessibility signals to deduce loop signals via a newly devised metric akin to the distortion in the Pearson correlation coefficient. Thus, InferLoop facilitates the inference of cell-type-specific higher-order chromatin contact maps.

In terms of benchmarks of overall pipelines to analyze single-cell chromatin accessibility, there are some reference papers that compare various tools for performance evaluation. For instance, to benchmark the performance of dimension reduction in scATAC, Zhang et al. (37) compared snapATAC2 with tools including ArchR, and Signac. For the performance of data integration in scHi-C, the same study (37) compared snapATAC2 with tools including Higashi. To benchmark the performance of inferring peak-level chromatin interactions and generating chromatin loop, Gao et al. (41) compared ChromaFold with tools including Cicero, ChromaFold, C.Origami.

Enhanced AI/ML models are available to infer higher-order chromatin contact maps utilizing diverse epigenomic data sets

Furthermore, Artificial intelligence (AI)/Machine learning (ML) models have been developed to predict chromatin contact maps using a variety of epigenomic profiles along with genome sequence data. Representative examples in this field include Epiphany (9), AkitaR (47), and C.Origami (10) (Fig. 1, Supplementary Table 2). Epiphany utilizes five epigenomic signals (DNaseI, CTCF, H3K27ac, H3K27me3, H3K4me3) to predict normalized Hi-C contact maps (9). This model incorporates bidirectional long short-term memory (LSTM) layers to capture long-range dependencies and may optionally integrate a generative adversarial network to enhance the accuracy of the contact maps. Additionally, Akita (48) was initially developed to predict 3D genome structure solely from DNA sequence data. Akita employs a CNN that converts input DNA sequences into predicted locus-specific genome folding. It analyzes 1 Mb of DNA sequence and predicts contact frequency maps for all pairs of 2 kb bins within that region. By adapting the Akita model to predict cell-type-specific chromatin contact frequencies, AkitaR establishes a novel deep learning framework

integrating genome sequences with genome-wide RNA-DNA interactions, thereby investigating the roles of chromatin-ssociated RNAs in genome folding (47). Conversely, C.Origami is a multimodal deep neural network that performs de novo prediction of cell-type-specific chromatin organization using DNA sequence and two cell-type-specific genomic features, namely CTCF binding and chromatin accessibility.

Most of these methods were developed to integrate various bulk epigenome datasets and DNA sequence information to infer higher-order chromatin contact maps. During the training of these machine learning models, bulk histone modification ChIP-seq and CTCF ChIP-seq data were integrated at the tissue level. Consequently, while feasible, it remains challenging to apply these models directly to predict single-cell chromatin contact maps, even when single-cell epigenome profiles are available for the target biological systems. In the future, it will be beneficial to rigorously explore the applicability of these machine learning-based methods to single-cell epigenome data, particularly with recent advancements in single-cell chromatin modification profiling techniques. For example, Wu et al. developed single-cell CUT&Tag analysis, which can profile histone modifications in singlecells, as demonstrated with H3K27me3 profiling in PBMCs and human glioblastoma samples (49). Additionally, CoBATCH is a high-throughput technology for single-cell epigenomic profiling that supports not only histone modifications but also non-histone chromatin-binding proteins, including transcription factors (50). With these newly developed techniques, one can either reuse previously trained models to infer cell-type-specific chromatin contact maps or retrain the models using single-cell level epigenome profiles and scHi-C profiles as the training sets to develop more accurate prediction models tailored for single-cell datasets.

Biological applications of above approaches

Single-cell informatics to infer higher-order chromatin contact maps offers new biological insights into cellular heterogeneity in 3D genome structures across various biological contexts, such as developmental landscapes, cell-type transitions during cancer progression, and drug treatment. These methods have been applied to diverse tissue types including cortex, hematopoietic systems, pancreas, muscle stem cells, and cancer samples.

For example, Zhou et al. (25) investigated the higher-order chromatin structure of breast cancer cells by integrating scHi-C and scRNA-seq data, and identified two cell subpopulations, known as topologically integrated subpopulations (TISPs), that closely resemble high cycling breast cancer persister cells. This cell population activates enzymes that increase chromatin interaction activities for cycling genes, leading to the emergence of reversible cancer persister cells upon drug treatment.

Chang et al. (51) developed Droplet Hi-C, utilizing a commercial microfluidic device for high-throughput, single-cell chromatin conformation profiling in droplets, to meticulously explore higher-order chromatin contact maps on a large scale. By implementing Droplet Hi-C on cortex tissues from 8-week-old mice and generating 6,235 single-cell chromatin profiles, researchers could examine cell-type specific 3D chromatin contact maps. Features of chromatin organization, such as compartments, domains, and loops, were comparable among different cell-types. Moreover, Droplet Hi-C was employed on two colorectal cancer cell lines to deduce structural variants and extrachromosomal DNA (ecDNA) from higher-order chromatin maps. Additionally, this technique uncovered ecDNA heterogeneity and evolutionary changes in glioblastoma samples, which were linked to the development of drug resistance following erlotinib treatment.

Pang et al. (52) investigated the 3D genome organization in the context of epithelial-mesenchymal transition across various cancer cell lines. To enhance the precision of comparing cells in the epithelial and mesenchymal states, this study adopted a novel method involving immunofluorescence staining for E and M markers (representing EpCAM and Vimentin, respectively) before isolating single cells using a dielectrophoretic platform, followed by single-cell Hi-C profiling. Notably, variability in the 3D genome structure of chromosome 2, where EPCAM is located, was observed among single cells in the E, hybrid, and M clusters.

Zhou et al. (24) developed GAGE-seq, which profiles multi-scale 3D genome organization and gene expression in single-cells. GAGE-seq concurrently profiles scHi-C and scRNA-seq from the same single-cell, with matched scHi-C and scRNA-seq profiles identified through well-specific barcoding combinations. Applied to cells from the mouse cortex aged 8-9 weeks, this technology generated 3,296 joint single-cell profiles of chromatin interactions and transcriptomes, revealing 28 known cell types across three major lineages in the mouse cortex. This data suggested that fine-scale CRE-chromatin looping, rather than alterations in large-scale 3D features, may drive the cell-type-specific expression of Epha4. Furthermore, this technology was employed on CD34+ cells from human bone marrow to investigate the dynamic relationship between 3D genome structure and gene expression during hematopoiesis.

Wang et al. (53) developed a deep-learning framework named EagleC to infer structural variants using higher-order chromatin contact maps from both bulk and single-cell Hi-C datasets. This framework was constructed utilizing a training set that included both Hi-C data and structural variant information obtained from WGS. They applied this novel model to identify structural variants in 105 cancer cell lines or primary tumors, predicting a total of 5620 SVs across these samples.

Gao et al. developed and applied ChromaFold (41) to their scATAC-seq datasets to infer higher-order chromatin contact maps in complex tissues. For example, ChromaFold resolved chromatin interactions in alpha and beta cells within the pancreatic islet cell populations derived from the scATAC and bulk Hi-C data of non-diabetic islet donors. In this sample, it visually depicted interactions at alpha and beta cell marker genes glucagon (GCG) and insulin (INS). The model predicted numerous contacts between the GCG gene and distal chro-

matin regions solely in alpha cells. Conversely, it indicated a heightened number of interactions between the INS gene and adjoining chromatin regions in beta cells compared to alpha cells, suggesting a cell-type specific configuration of higher-order chromatin maps.

Yang et al. explored 3D chromatin reorganisation in the context of muscle stem cell aging (54). In this study, the authors compared the higher-order chromatin contact maps of young and aged mice muscle stem cells using bulk Hi-C and scATAC-seq. To analyze bulk Hi-C they utilized HiCExplorer (17) and HiCCUPS algorithms (55). To reconstruct chromatin loops based on scATAC-seq, they utilized ArchR and Cicero tools.

DISCUSSION

Pros and cons of building chromatin contact maps using scHi-C and scATAC-seq

To date, we have reviewed the single-cell bioinformatics methods for constructing chromatin contact maps using single-cell epigenomic datasets, such as scHi-C and scATAC-seq. scHi-C based methods offer the advantage of directly measuring chromatin interactions between two genomic regions, though the data may be sparse due to the single-cell nature. However, numerous experimental protocols exist for generating scHi-C data, each necessitating specialized bioinformatics tools for processing. Consequently, it is challenging to establish a definitive one-step pipeline for analyzing scHi-C data. In contrast, the experimental protocol for generating scATAC-seq data is relatively standardized, primarily using the 10X genomics kit for scATAC-seq library production. This standardization benefits the bioinformatics pipeline for scATAC-seq analysis, facilitating meta analyses that integrate datasets from different laboratories or resources. Although higher-order chromatin contact maps derived from scATAC-seq result from inference rather than direct measurement, they offer the opportunity to integrate and reanalyze existing scATAC-seq data, thereby deriving novel biological hypotheses from different analytical perspectives.

Can long-read sequencing be helpful in generating higher-order contact maps?

In the preceding sections, we primarily explored single-cell bioinformatics methods for analyzing short-read sequencing datasets. Conversely, long-read sequencing technologies have recently become more accessible to researchers due to technical advancements and enhanced sequencing accuracy. These technologies are advantageous for analyzing repetitive and complex genomic regions such as centromeres, detecting long-range structural rearrangements, and accurately phasing haplotypes (56). Long-read sequencing methodologies have also been employed in single-cell Hi-C (20) and single-cell ATAC-seq (57) studies. In conventional scHi-C methods, the limitations on read length made it difficult to detect multiway interactions involving multiple (more than three) regulatory elements or genomic regions in single cells. As an alternative, the long-read single-cell Hi-C approach, such as scNanoHi-C, has proven effective in elucidating multiway higher-order interactions. Similarly, a multiway contact map can aid in the detection of complex chromosomal rearrangements involving at least three breakpoints across two or more chromosomes (58). Long-read sequencing is also beneficial for phasing haplotypes in both single-cell Hi-C and single-cell ATAC-seq sequencing reads. This facilitates the construction of a haplotype-resolved higher-order chromatin contact map and the identification of allele-specific chromatin interaction patterns based on scHi-C and scATAC-seq analysis.

Can AI/ML models be helpful for generating higher-order contact maps?

In this review, we have summarized AI/ML models designed for generating higher-order contact maps through DNA sequence analysis and integrating histone modification and chromatin accessibility data. However, most AI/ML models for this purpose have been developed to utilize bulk epigenomic datasets rather than single-cell datasets. Consequently, there is a need either to adapt these models for single-cell epigenomic data or to develop new AI/ML models specifically for single-cell epigenomic datasets including scATAC-seq. Moreover, with the progression of long-read sequencing and haplotype-phasing techniques, AI/ML models could be newly trained to infer haplotype-specific chromatin contact maps in the future.

Further applications: single-cell contact maps using scMNase-seq, Strand-seq, or spatial epigenetic methods

In this review, we surveyed and summarized the methods used to generate higher-order chromatin contact maps, primarily through scHi-C and scATAC-seq. Nevertheless, various other single-cell technologies capable of measuring different aspects of single-cell epigenomes, such as scMNase-seq (59), and Strand-seq (60), may also provide valuable inputs for inferring higher-order chromatin contact maps, assuming appropriate bioinformatics methods are developed. The inferred higher-order chromatin contact maps could then be integrated with newer modalities, such as nucleosome occupancy (scMNase-seq) and structural variants (Strand-seq). Moreover, recent advancements in Spatial omics technology allow profiling of single-cell epigenomes in a spatially resolved manner (61-63). To our knowledge, until the completion of this mini-review, no publications have investigated spatially-resolved higher-order chromatin contact maps in tissues. Thus, in the future, the active development of novel bioinformatics methods will be essential for inferring and examining spatially resolved higher-order chromatin maps.

ACKNOWLEDGEMENTS

This work was supported by NRF (RS-2024-00336518, NRF-2018R1A6A1A03025607), POSCO Science Fellowship of POSCO

TJ Park Foundation 2024-11-1819, and the Yonsei University Research Fund of 2024-22-0044.

CONFLICTS OF INTEREST

The authors have no conflicting interests.

REFERENCES

1. Jung N and Kim TK (2021) Advances in higher-order chromatin architecture: the move towards 4D genome. BMB Rep 54, 233-245
2. Harris HL, Gu H, Olshansky M et al (2023) Chromatin alternates between A and B compartments at kilobase scale for subgenic organization. Nat Commun 14, 3303
3. Xiong K and Ma J (2019) Revealing Hi-C subcompartments by imputing inter-chromosomal chromatin interactions. Nat Commun 10, 5069
4. Cosenza MR, Rodriguez-Martin B and Korbel JO (2022) Structural variation in cancer: role, prevalence, and mechanisms. Annu Rev Genomics Hum Genet 23, 123-152
5. Krupina K, Goginashvili A and Cleveland DW (2024) Scrambling the genome in cancer: causes and consequences of complex chromosome rearrangements. Nat Rev Genet 25, 196-210
6. Pitel BA, Zuckerman EZ and Baughn LB (2023) Mate pair sequencing: next-generation sequencing for structural variant detection. Methods Mol Biol 2621, 127-149
7. Song F, Xu J, Dixon J and Yue F (2022) Analysis of Hi-C data for discovery of structural variations in cancer. Methods Mol Biol 2301, 143-161
8. Fang K, Wang J, Liu L and Jin VX (2022) Mapping nucleosome and chromatin architectures: a survey of computational methods. Comput Struct Biotechnol J 20, 3955-3962
9. Yang R, Das A, Gao VR et al (2023) Epiphany: predicting Hi-C contact maps from 1D epigenomic signals. Genome Biol 24, 134
10. Tan J, Shenker-Tauris N, Rodriguez-Hernaez J et al (2023) Cell-type-specific prediction of 3D chromatin organization enables high-throughput *in silico* genetic screening. Nat Biotechnol 41, 1140-1150
11. Clark SJ, Lee HJ, Smallwood SA, Kelsey G and Reik W (2016) Single-cell epigenomics: powerful new methods for understanding gene regulation and cell identity. Genome Biol 17, 72
12. Heumos L, Schaar AC, Lance C et al (2023) Best practices for single-cell analysis across modalities. Nat Rev Genet 24, 550-572
13. Xing Z, Mai H, Liu X et al (2022) Single-cell diploid Hi-C reveals the role of spatial aggregations in complex rearrangements and KMT2A fusions in leukemia. Genome Biol 23, 173
14. Ward CM, To TH and Pederson SM (2020) ngsReports: a Bioconductor package for managing FastQC reports and other NGS related log files. Bioinformatics 36, 2587-2588
15. Li H and Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754-1760
16. Ramani V, Deng X, Qiu R et al (2017) Massively multiplex single-cell Hi-C. Nat Methods 14, 263-266
17. Wolff J, Rabbani L, Gilsbach R et al (2020) Galaxy HiCExplorer 3: a web server for reproducible Hi-C, capture Hi-C and single-cell Hi-C data analysis, quality control and visualization. Nucleic Acids Res 48, W177-W184
18. Wang Y and Cheng J (2024) HiCDiff: single-cell Hi-C data denoising with diffusion models. Brief Bioinform 25
19. Open2C, Abdennur N, Fudenberg G et al (2024) Pairtools: from sequencing data to chromosome contacts. PLoS Comput Biol 20, e1012164
20. Li W, Lu J, Lu P et al (2023) scNanoHi-C: a single-cell long-read concatemer sequencing method to reveal high-order chromatin structures within individual cells. Nat Methods 20, 1493-1505
21. Tan L, Xing D, Chang CH, Li H and Xie XS (2018) Three-dimensional genome structures of single diploid human cells. Science 361, 924-928
22. Ye Y, Gao L and Zhang S (2019) Circular trajectory reconstruction uncovers cell-cycle progression and regulatory dynamics from single-cell Hi-C Maps. Adv Sci (Weinh) 6, 1900986
23. Wu Y, Shi Z, Zhou X et al (2024) scHiCyclePred: a deep learning framework for predicting cell cycle phases from single-cell Hi-C data using multi-scale interaction information. Commun Biol 7, 923
24. Zhou T, Zhang R, Jia D et al (2024) GAGE-seq concurrently profiles multiscale 3D genome organization and gene expression in single cells. Nat Genet 56, 1701-1711
25. Zhou Y, Li T, Choppavarapu L, Fang K, Lin S and Jin VX (2024) Integration of scHi-C and scRNA-seq data defines distinct 3D-regulated and biological-context dependent cell subpopulations. Nat Commun 15, 8310
26. Liu Z, Chen Y, Xia Q et al (2023) Linking genome structures to functions by simultaneous single-cell Hi-C and RNA-seq. Science 380, 1070-1076
27. Yu M, Abnousi A, Zhang Y et al (2021) SnapHiC: a computational pipeline to identify chromatin loops from single-cell Hi-C data. Nat Method 18, 1056-1059
28. Lee L, Yu M, Li X et al (2023) SnapHiC-D: a computational pipeline to identify differential chromatin contacts from single-cell Hi-C data. Brief Bioinform 24
29. Liu W, Zhong W, Giusti-Rodriguez P et al (2024) SnapHiC-G: identifying long-range enhancer-promoter interactions from single-cell Hi-C data via a global background model. Brief Bioinform 25, 1-14
30. Xiong K, Zhang R and Ma J (2024) scGHOST: identifying single-cell 3D genome subcompartments. Nat Methods 21, 814-822
31. Zhang R, Zhou T and Ma J (2022) Multiscale and integrative single-cell Hi-C analysis with Higashi. Nat Biotechnol 40, 254-261
32. Hua D, Gu M, Zhang X et al (2024) DiffDomain enables identification of structurally reorganized topologically associating domains. Nat Commun 15, 502
33. Flyamer IM, Illingworth RS and Bickmore WA (2020) Coolpup.py: versatile pile-up analysis of Hi-C data. Bioinformatics 36, 2980-2985
34. Baek S and Lee I (2020) Single-cell ATAC sequencing

analysis: from data preprocessing to hypothesis generation. Comput Struct Biotechnol J 18, 1429-1439
35. Satpathy AT, Granja JM, Yost KE et al (2019) Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. Nat Biotechnol 37, 925-936
36. Tayyebi Z, Pine AR and Leslie CS (2024) Scalable and unbiased sequence-informed embedding of single-cell ATAC-seq data with CellSpace. Nat Methods 21, 1014-1022
37. Zhang K, Zemke NR, Armand EJ and Ren B (2024) A fast, scalable and versatile tool for analysis of single-cell omics data. Nat Methods 21, 217-227
38. Stuart T, Srivastava A, Madad S, Lareau CA and Satija R (2021) Single-cell chromatin state analysis with Signac. Nat Methods 18, 1333-1341
39. Granja JM, Corces MR, Pierce SE et al (2021) ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. Nat Genet 53, 403-411
40. Yu W, Uzun Y, Zhu Q, Chen C and Tan K (2020) scATAC-pro: a comprehensive workbench for single-cell chromatin accessibility sequencing data. Genome Biol 21, 94
41. Gao VR, Yang R, Das A et al (2024) ChromaFold predicts the 3D contact map from single-cell chromatin accessibility. Nat Commun 15, 9432
42. Fortin JP and Hansen KD (2015) Reconstructing A/B compartments as revealed by Hi-C using long-range correlations in epigenetic data. Genome Biol 16, 180
43. Buenrostro JD, Wu B, Litzenburger UM et al (2015) Single-cell chromatin accessibility reveals principles of regulatory variation. Nature 523, 486-490
44. Pliner HA, Packer JS, McFaline-Figueroa JL et al (2018) Cicero predicts cis-regulatory DNA interactions from single-cell chromatin accessibility data. Mol Cell 71, 858-871.e858
45. Bravo Gonzalez-Blas C, De Winter S, Hulselmans G et al (2023) SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. Nat Method 20, 1355-1367
46. Zhang F, Jiao H, Wang Y et al (2023) InferLoop: leveraging single-cell chromatin accessibility for the signal of chromatin loop. Brief Bioinform 24, 1-15
47. Kuang S and Pollard KS (2024) Exploring the roles of RNAs in chromatin architecture using deep learning. Nat Commun 15, 6373
48. Fudenberg G, Kelley DR and Pollard KS (2020) Predicting 3D genome folding from DNA sequence with Akita. Nat Methods 17, 1111-1117
49. Wu SJ, Furlan SN, Mihalas AB et al (2021) Single-cell CUT&Tag analysis of chromatin modifications in differentiation and tumor progression. Nat Biotechnol 39, 819-824
50. Wang Q, Xiong H, Ai S et al (2019) CoBATCH for high-throughput single-cell epigenomic profiling. Mol Cell 76, 206-216.e207
51. Chang L, Xie Y, Taylor B et al (2024) Droplet Hi-C enables scalable, single-cell profiling of chromatin architecture in heterogeneous tissues. Nat Biotechnol [Online ahead of print]
52. Pang QY, Tan TZ, Sundararajan V et al (2022) 3D genome organization in the epithelial-mesenchymal transition spectrum. Genome Biol 23, 121
53. Wang X, Luan Y and Yue F (2022) EagleC: a deep-learning framework for detecting a full range of structural variations from bulk and single-cell contact maps. Sci Adv 8, eabn9215
54. Yang BA, Larouche JA, Sabin KM, Fraczek PM, Parker SCJ and Aguilar CA (2023) Three-dimensional chromatin re-organization during muscle stem cell aging. Aging Cell 22, e13789
55. Rao SS, Huntley MH, Durand NC et al (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 159, 1665-1680
56. Marx V (2023) Method of the year: long-read sequencing. Nat Methods 20, 6-11
57. Hu Y, Jiang Z, Chen K et al (2023) scNanoATAC-seq: a long-read single-cell ATAC sequencing method to detect chromatin accessibility and genetic variants simultaneously within an individual cell. Cell Res 33, 83-86
58. Pellestor F, Anahory T, Lefort G et al (2011) Complex chromosomal rearrangements: origin and meiotic behavior. Hum Reprod Update 17, 476-494
59. Lai B, Gao W, Cui K et al (2018) Principles of nucleosome organization revealed by single-cell micrococcal nuclease sequencing. Nature 562, 281-285
60. Jeong H, Grimes K, Rauwolf KK et al (2023) Functional analysis of structural variants in single cells using Strand-seq. Nat Biotechnol 41, 832-844
61. Zhang D, Deng Y, Kukanja P et al (2023) Spatial epigenome-transcriptome co-profiling of mammalian tissues. Nature 616, 113-122
62. Lu T, Ang CE and Zhuang X (2022) Spatially resolved epigenomic profiling of single cells in complex tissues. Cell 185, 4448-4464.e4417
63. Nikopoulou C, Kleinenkuhnen N, Parekh S et al (2023) Spatial and single-cell profiling of the metabolome, transcriptome and epigenome of the aging mouse liver. Nat Aging 3, 1430-1445