Journal of Computer Science and Artificial Intelligence
ISSN: 3078-8242 (Print), ISSN: 3078-2309 (Online) | Vol. 7, No. 1, 2026

Research Progress and Frontier Prospects of Chromatin A/B Compartment Classification Methods

Xiangchao Meng *

Henan Polytechnic University, School of Physics & Electronic Information Engineering, Jiaozuo 454000, China
957384974@qq.com

Abstract: Accurate classification of chromatin A/B compartments is one of the core issues in three-dimensional genomics, which is critical for understanding how genome spatial organization affects gene regulation, cell fate determination, and disease pathogenesis. Since Hi-C technology first revealed in 2009 that the genome is spatially segregated into transcriptionally active A compartments and inactive B compartments, principal component analysis (PCA) based on Hi-C contact matrices has long served as the standard strategy for compartment identification. However, with the in-depth advancement of research toward single-cell resolution, cross-cell-type prediction, and the integration of multimodal epigenomic data, the inherent limitations of conventional PCA have become increasingly prominent. In recent years, the introduction of machine learning and deep learning has brought paradigm innovation to chromatin compartment classification. Methodological evolution ranges from early stacked artificial neural networks based on sequence-derived features, to convolutional neural networks utilizing raw DNA sequences, and further to advanced approaches integrating recurrent neural networks, Transformer architectures, graph-theoretic optimization, and interpretable learning, which have substantially expanded the theoretical framework and application scope of compartment analysis. This review systematically summarizes the research progress of chromatin A/B compartment classification methods and covers more than ten representative approaches, including SACSANN, ABCNet, CoRNN, TECSAS, MaxComp, DeepExDC, HiC-SCA, ABCRNet, SCI and CDACHIE. We comprehensively summarize and comment on existing advances from three dimensions: biological principles and experimental detection techniques, machine learning-based methods, and deep learning-driven strategies. Furthermore, we discuss the major challenges currently restricting this field, such as limited cross-cell-type generalization, insufficient model interpretability, the absence of unified benchmarking criteria, and widespread spatial heterogeneity in compartment annotation. Finally, we highlight future research directions, including multimodal data integration, cross-domain adaptation of foundation models, and the translational application of compartment prediction tools to clinical research scenarios such as cancer genomics and developmental biology.

Keywords: Chromatin A/B compartments, Machine learning, Deep learning, 3D genome organization.

1. Introduction

The three-dimensional spatial structure of chromatin within the nucleus of eukaryotic cells serves as a core determinant of genome function, directly regulating gene transcription, the timing of DNA replication, and the stable maintenance of cellular identity. In 2009, the pioneering Hi-C study conducted by the Lieberman-Aiden team demonstrated that the genome is spatially partitioned into two distinct compartments, namely A and B. Compartment A is gene-dense and transcriptionally active, whereas compartment B is gene-sparse and associated with transcriptionally silent and repressed chromatin states. Since this landmark discovery, compartment annotation has become an indispensable initial step in nearly all three-dimensional genomics workflows. It facilitates the interpretation of long-range chromatin interactions and enables the characterization of large-scale genome structural reorganization during cell differentiation, disease progression and species evolution. For more than a decade, principal component analysis (PCA) based on normalized Hi-C contact matrices has remained the gold standard for chromatin compartment identification. This method determines compartment types according to the sign and magnitude of the first eigenvector, featuring low computational cost and clear mechanistic principles [1].

Nevertheless, with the advancement of high-resolution profiling, single-cell sequencing and multi-omics integration research, the inherent limitations of PCA have become increasingly prominent. First, PCA only captures population-averaged chromatin features and masks widespread cell-to-cell compartment heterogeneity, which has been well documented by single-cell imaging and single-cell Hi-C studies. Second, genomic characteristics and spatial localization exhibit highly nonlinear correlations, while PCA adopts linear decomposition for modeling, which fails to conform to genuine biological mechanisms. Third, PCA is heavily reliant on high-quality Hi-C data, which suffer from high experimental costs and complicated library construction procedures. Such drawbacks greatly restrict its application in rare cell populations, clinical specimens and non-model organisms [2]. To address these bottlenecks, researchers have developed novel computational approaches to infer compartment profiles using readily accessible genomic and epigenomic features, or to directly extract compartment information from emerging data modalities, such as single-cell 3D structural coordinates and chromatin tracing microscopy data.

Machine learning and deep learning excel at capturing complex nonlinear relationships among DNA sequences, epigenomic signatures and 3D chromatin architecture, and have thus become the core technical foundation for next-generation compartment analytical tools. In the past five years, neural network architectures have undergone rapid evolution, ranging from early convolutional neural networks trained on raw genomic sequences to advanced Transformer models capable of integrating numerous epigenomic signals across

megabase-scale genomic regions. Meanwhile, the scope of compartment analysis has expanded far beyond the binary A/B classification within a single cell type, covering fine-scale subcompartment annotation, cross-species predictive transfer, unsupervised structural identification from 3D genome conformations, and statistically rigorous differential analysis at single-cell resolution. Amid rapid technological innovation, there is an urgent demand to systematically summarize the core principles and biological implications of existing algorithms. Previous reviews have merely regarded compartment prediction as a peripheral topic within 3D genome bioinformatics, while few studies have specifically focused on the design philosophy, input data strategies and translational potential of current machine learning-based methods for A/B compartment classification.

2. Biological Basis and Experimental Detection Technologies of Chromatin A/B Compartments

2.1. Origin and Biological Connotation of Chromatin Compartmentalization

The three-dimensional spatial organization of eukaryotic genomes within the nucleus is randomly distributed but follows strict hierarchical structural principles, which are tightly coupled with genome functions, including gene regulation, DNA replication, and cell differentiation. Within the hierarchical architecture of the 3D genome, chromatin compartments represent one of the earliest identified and most fundamental organizational layers. The core connotation of the compartment concept lies in that genomic regions with similar transcriptional activity and chromatin states tend to spatially cluster and preferentially interact with one another, thereby forming functionally distinct and spatially segregated domains.

In 2009, Lieberman-Aiden and colleagues first observed a large-scale plaid-like interaction pattern in Hi-C contact maps, revealing that the genome is spatially partitioned into at least two major compartments: gene-dense, transcriptionally active A compartments, and gene-poor, transcriptionally silent B compartments [3]. Such spatial segregation is highly correlated with the biochemical properties of chromatin. Specifically, A compartments correspond to euchromatin enriched with active histone modifications such as H3K27ac and H3K4me3, and are generally localized in the nuclear interior. In contrast, B compartments represent heterochromatin marked by repressive epigenetic signatures including H3K9me3, and are frequently anchored to the nuclear lamina.

2.2. Hierarchical Organization and Subcompartments of Compartments

With the continuous improvement of Hi-C data resolution and sequencing depth, researchers have found that the traditional binary A/B compartment framework is insufficient to fully characterize the complexity of compartmental organization. In 2014, Rao et al. [4] systematically identified six subcompartments (A1, A2, B1, B2, B3, and B4) using high-resolution Hi-C data, each exhibiting distinct chromatin interaction patterns and functional features. H3K27ac performs well in discriminating A/B compartments and is recognized as one of the crucial predictive epigenetic markers.

The discovery of subcompartments has greatly advanced our understanding of chromatin spatial organization, indicating that even within the same major compartment, fine-grained differences exist in functional states, nuclear localization, and interaction preferences. In recent years, the TECSAS method enables the direct prediction of five major subcompartments, including A1, A2, B1, B2 and B3. Meanwhile, the CDACHIE approach identifies six functional domains through contrastive learning, partially revealing potential functional subtypes that can be further divided within the conventional A2 compartment.

2.3. Experimental Detection Technologies

Experimental detection of chromatin compartments relies on chromatin conformation capture techniques and their derived methods. Among them, Hi-C has become the gold standard for compartment detection by quantitatively measuring the contact frequency between chromatin loci across the whole genome. The typical experimental and analytical workflow includes crosslinking to fix chromatin conformation, restriction enzyme digestion, biotin-marked ligation, purification and high-throughput sequencing, ultimately generating a genome-wide contact matrix. After matrix balancing and distance normalization of the contact matrix, the Pearson correlation coefficient matrix is calculated. Principal component analysis is further adopted to extract eigenvectors for compartment identification, in which the positive and negative values of PC1 generally correspond to B and A compartments, respectively, with direction correction implemented according to genomic features such as gene density and GC content. In addition, the emerging Micro-C technology can provide nucleosome-level ultra-high-resolution contact maps, while multiplexed FISH imaging enables direct visualization of 3D chromatin configurations in single cells, offering complementary experimental approaches for compartment analysis.

2.4. Limitations of Traditional Computational Methods and Driving Forces of Computational Innovation

Although PCA-based compartment identification methods have been widely applied for more than a decade, their inherent limitations have become increasingly prominent. PCA relies on population-averaged Hi-C contact matrices and fails to resolve cell-to-cell compartment heterogeneity at the single-cell level. Its linear decomposition assumption also struggles to capture the complex nonlinear relationships between genomic features and spatial localization. More importantly, the high cost and technical barriers of Hi-C data acquisition pose critical constraints, as high-quality Hi-C data are particularly difficult to obtain for rare cell types, clinical specimens and non-model species.

These limitations have strongly driven computational methodological innovations and prompted researchers to explore two major technical routes. The first strategy infers compartment identities using readily accessible genomic and epigenomic profiles, thereby circumventing the direct dependence on Hi-C experiments. The second direction focuses on developing novel algorithmic frameworks to directly extract compartment signatures from emerging single-cell datasets, such as single-cell Hi-C and three-dimensional imaging data. Under such circumstances, machine learning and deep learning have been systematically introduced into the field of chromatin compartment classification.

3. Research on Chromatin A/B Compartment Classification Based on Machine Learning

The application of machine learning methods in the field of chromatin compartment classification marks a critical shift of compartment analysis from purely descriptive statistics to predictive modeling. Compared with deep learning, the approaches summarized in this section mainly adopt strategies such as shallow network architectures, traditional feature engineering and graph optimization, featuring high computational efficiency, strong interpretability and low data requirements.

SACSANN [5] (Stacked Artificial Neural Networks for Chromatin Structural Annotations) was proposed by Prost, Cameron and Blanchette in 2020, serving as one of the early machine learning-based methods to predict chromatin compartments solely from DNA sequences. Instead of directly learning patterns from raw DNA sequences, its core strategy relies on elaborate feature engineering to extract high-level genomic features, which are subsequently fed into stacked artificial neural networks for compartment prediction.

Specifically, the sequence-derived features adopted by SACSANN include: (1) the distribution density of predicted transcription factor binding sites across genomic regions; (2) the compositional profiles and distribution patterns of transposable elements, such as LINEs, SINEs and LTRs; (3) additional sequence-derived signatures closely associated with genomic functions. These features are aggregated within uniformly sized genomic bins (typically 100 kb) and then utilized to train a multi-layer stacked artificial neural network.

The major contributions of SACSANN are reflected in three aspects. First, it enables accurate cell-type-specific compartment prediction. The model generates context-dependent compartment annotations across distinct cell types, rather than producing fixed sequence-determined results. Second, it exhibits prominent cross-species transferability between human and mouse homologous cell types, indicating that the sequence-based determinants of chromatin compartmentalization are partially evolutionarily conserved across mammals. Third, through feature importance analysis, SACSANN can pinpoint critical genomic sequence drivers that dominate A/B compartment identity, providing mechanistic insights into the evolutionary principles of 3D genome organization. Moreover, the original source code of SACSANN has been openly released on GitHub, offering a feasible computational tool for 3D genome research in non-model organisms lacking high-quality Hi-C datasets.

HiC-SCA [6] (HiC Spectral Compartment Assignment) was proposed by Chan and Kono in 2025. It reformulates the task of identifying chromatin compartments from Hi-C contact matrices as a graph-based machine learning problem. This method models chromatin as a weighted network, in which nodes represent genomic bins and edge weights between nodes encode Hi-C interaction frequencies. Based on this graph representation, HiC-SCA employs spectral clustering to optimally partition the nodes of each chromosome into two groups corresponding to compartments A and B. This strategy effectively avoids the eigenvector sign ambiguity and subjective resolution selection inherent to conventional PCA-based pipelines. HiC-SCA incorporates three key technical improvements: (1) a noise filtering module that removes low-quality signals to prevent erroneous compartment assignment; (2) an automated orientation assignment strategy that accurately distinguishes compartment A from compartment B without relying on external genomic annotations; (3) a quantitative quality scoring system coupled with a data-driven resolution selection scheme, which evaluates the reliability of compartment calling and determines the highest analyzable resolution directly from the intrinsic properties of Hi-C data. Evaluations across 21 Hi-C datasets covering diverse cell types demonstrate that HiC-SCA outperforms existing methods in cross-dataset consistency. It yields highly reproducible compartment annotations for independent datasets of the same cell type, conferring unique practical advantages for differential analysis across multiple experimental conditions and biological replicates.

Proposed by Gill in 2025, ABCRNet [7] is a CNN-based approach designed to predict the Compartment Membership Ratio (CMR) of A/B compartments for each genomic bin from reference genome sequences. Most previous compartment prediction studies simply assigned each bin a discrete label of "A" or "B", ignoring widespread ambiguous compartment identities across cell types. Specifically, a given bin may adopt an A-compartment state in certain cell types while shifting to a B-compartment state in others. Distinct from traditional hard classification strategies, the core innovation of ABCRNet lies in quantifying the compartment preference of each bin using a continuous value ranging from 0 to 1. The model takes raw DNA sequences of the reference genome as input. Training labels are generated by calculating the frequency ratio of A-compartment occupancy for each bin across multiple human and mouse Hi-C datasets covering diverse cell types. Leveraging convolutional neural networks, ABCRNet automatically learns the mapping between sequence signatures and the variability of A/B compartment membership. Validations show that ABCRNet achieves robust predictive performance in genomic regions with extremely high or low GC content, demonstrating that GC composition serves as a pivotal sequence determinant for compartment membership estimation. This study uncovers inherent correlations between sequence composition and cell-type-specific compartment variability, which can be effectively captured by deep learning. It provides novel insights into how primary genome sequences encode the structural diversity of chromatin compartments within cell populations.

MaxComp [8] is an unsupervised method formally published in PLOS Computational Biology by Zhan et al. in 2025. From a methodological perspective, it adopts graph-theoretic optimization rather than the conventional supervised learning paradigm and is therefore discussed in this chapter. The core innovation of MaxComp lies in fundamentally reformulating single-cell chromatin compartment prediction as a max-cut problem in graph theory. Its fundamental principle is that genomic regions within the same compartment are spatially proximal and should thus be classified into the same group in the structural graph, whereas regions belonging to different compartments should be separated on opposite sides of the cut.

The detailed technical pipeline is as follows: (1) The single-cell three-dimensional structure of each chromosome is constructed as an undirected graph, where nodes represent chromatin loci and edge weights encode spatial distances or structural information derived from 3D coordinates; (2) Compartment prediction is transformed into searching for an optimal bipartition of the graph that maximizes the total

weight of edges crossing the cut. This principle is fully consistent with the geometric intuition that loci from distinct compartments tend to be spatially distant; (3) Semidefinite programming is employed to solve the max-cut problem, yielding the optimal binary assignment for each node, which corresponds to single-cell A/B compartment annotations. The data sources of MaxComp include direct 3D spatial coordinates obtained from multiplexed FISH imaging and 3D structural models reconstructed from Hi-C data. The major findings of MaxComp are summarized as follows: (1) Population-averaged single-cell compartment annotations generated by MaxComp show high consistency with PCA results from bulk Hi-C data, demonstrating that compartmentalization patterns can be restored relying solely on geometric principles; (2) This method reveals prevalent cell-to-cell compartment heterogeneity, in which identical genomic loci exhibit divergent compartment identities across individual cells; (3) It establishes a direct link between single-cell chromosomal spatial organization and transcriptional activity based on multiplexed FISH imaging datasets.

Notably, several alternative approaches adopt graph embedding or unsupervised learning strategies for subcompartment prediction. SCI [9] (Sub-Compartment Identifier) integrates graph embedding with unsupervised learning to infer subcompartments from Hi-C chromatin interaction data, and achieves superior performance in network topological centrality and clustering metrics compared with hidden Markov model-based methods. CDACHIE (Chromatin Domain Annotation by Integrating Chromatin Interaction and Epigenomic Data with Contrastive Learning) was published in Bioinformatics by Yoshinaga and Maruyama in 2025. This method integrates Hi-C structural profiles with twelve epigenetic signatures (including H3K27ac, H3K4me3 and other histone marks) within a contrastive learning framework. It employs a bidirectional encoder to produce aligned embedding vectors and adopts K-means clustering to identify six distinct functional chromatin domains. In the GM12878 cell line, CDACHIE outperforms HMM_combined and GMM_GBR in multiple evaluation indicators, such as the explained variance of replication timing and the observed-to-expected ratio of CTCF ChIA-PET loops. These results highlight the unique advantages of contrastive learning in bridging the resolution gap between Hi-C interactome and epigenomic datasets.

4. Research on Chromatin A/B Compartment Classification Based on Deep Learning

Deep learning methods, with their capabilities of end-to-end feature learning, multi-layer nonlinear transformation, and large-scale multimodal data modeling, have opened new avenues for chromatin compartment classification research. This section focuses on representative approaches based on deep neural networks and systematically reviews them by architectural categories.

ABCNet [10] is a convolutional neural network proposed by Kirchhof in his master's thesis in 2021, and it stands as one of the earliest compartment prediction models in this field that completely bypasses manual feature engineering. Its design philosophy is concise yet profound: given that compartment identity is closely associated with chromatin activity, and such activity is ultimately encoded by genomic DNA sequences—through sequence-dependent molecular events including transcription factor binding and histone modification—whether sufficient intrinsic information exists to directly infer compartment status from raw DNA sequences alone. To verify this hypothesis, ABCNet takes reference genome sequences as the sole input. The genome is partitioned into uniformly sized bins, typically at a 100 kb resolution. The sequence of each bin is converted into one-hot encoding and fed into a CNN architecture consisting of one-dimensional convolutional layers, pooling layers, and fully connected layers, which ultimately outputs the predicted probability of each bin belonging to the A or B compartment.

Experimental results demonstrate that relying solely on DNA sequences, ABCNet achieves competitive accuracy comparable to conventional methods dependent on manually engineered features. This strongly validates that genomic sequences inherently contain sufficient information to reconstruct three-dimensional compartmentalization patterns. Furthermore, latent space analysis of ABCNet reveals key sequence signatures linked to higher-order genome organization, providing interpretable clues for understanding how primary sequences encode layered chromatin structures. A major limitation of ABCNet lies in its exclusive dependence on static DNA sequences, which restricts its capacity to capture cell-type-specific epigenetic modifications and dynamic regulatory features.

CoRNN [11] (Compartment prediction using Recurrent Neural Networks) was published in iScience in 2024 by Zheng, Thakkar, Harris and colleagues. It is a deep learning tool for A/B compartment prediction based on histone modification enrichment data, with a core advantage of robust cross-cell-type generalization. Unlike ABCNet, which relies solely on DNA sequences, CoRNN takes ChIP-seq signals of histone modifications as inputs, fully leveraging the biological fact that histone marks serve as direct molecular indicators of chromatin activity states. The model adopts a recurrent neural network (RNN) architecture. Given the linear nature of chromosomes, histone modification patterns in adjacent genomic regions exhibit inherent sequential dependencies, and RNNs are uniquely suited to model such sequential correlations. After adequate training, CoRNN can be effectively transferred across cell types. It achieves an average AuROC of 90.9%, substantially outperforming traditional PCA-based pipelines and sequence-only prediction methods. In terms of feature importance analysis, CoRNN identifies H3K27ac (a hallmark of active enhancers and promoters) and H3K36me3 (a marker of transcriptional elongation) as the most predictive histone modifications. This evidence indicates that transcription-associated histone signatures act as the dominant epigenetic determinants governing the spatial positioning of chromatin compartments, providing mechanistic insights into the preferential localization of active euchromatin in the nuclear interior.

TECSAS [12] (Transformer of Epigenetics to Chromatin Structural Annotations) was first released on bioRxiv by Dodero-Rojas, Contessoto, Onuchic and co-workers in 2024 and formally published in PLOS Computational Biology in 2025. It represents the first method in this field to introduce the Transformer architecture into chromatin structural prediction. Equipped with multi-head self-attention mechanisms, the Transformer is inherently adept at capturing long-range dependencies across arbitrary genomic distances, which is particularly valuable for compartment prediction. Chromatin 3D interactions occur not only between neighboring regions but also through long-range contacts

spanning dozens of megabases. The input data of TECSAS integrates multiple histone modification ChIP-seq signals, chromatin accessibility profiles, transcription factor binding peaks, and RNA-Seq expression data. It outputs fine-scale subcompartment annotations (A1, A2, B1, B2, B3) at a typical resolution of 25–50 kb. With self-attention layers explicitly modeling long-range epigenetic dependencies, its core concept "You Only Need Epigenetics" delivers a critical biological conclusion: comprehensive, high-quality epigenomic data alone can precisely decode 3D chromatin compartmentalization, enabling accurate annotation even in emerging cell types without available Hi-C resources. Experimental results demonstrate that TECSAS achieves high accuracy in subcompartment classification. It highlights the essential contribution of long-range epigenomic context to genome folding and exhibits promising performance in predicting genomic loci associated with nuclear bodies, including the nuclear lamina, nucleoli, and nuclear speckles, thereby extending the functional scope of conventional compartment prediction tools. In terms of limitations, TECSAS requires a broad panel of epigenomic inputs, leading to high experimental costs for data generation. Meanwhile, the Transformer framework contains massive parameters, which imposes substantial computational demands for both model training and inference.

DeepExDC [13] was published in Briefings in Bioinformatics in 2025 by Lyu, Cao, Long and colleagues. It is the first interpretable deep learning method specifically designed for differential compartment analysis based on single-cell Hi-C data, filling a critical research gap in this subfield. Although single-cell Hi-C technology enables the profiling of higher-order chromatin structures at the individual cell level, the extreme sparsity and high noise inherent to scHi-C data severely limit the feasibility and performance of conventional PCA-based pipelines. More importantly, few prior approaches support statistically rigorous differential compartment analysis at the single-cell resolution. The analytical pipeline of DeepExDC consists of three core steps: (1) A one-dimensional convolutional neural network is adopted to learn compact compartmental feature representations for each genomic bin from sparse scHi-C contact maps; (2) The SHAP interpretability framework is integrated to quantify and biologically interpret the feature contribution of each genomic bin to compartment classification decisions; (3) Pairwise distance matrices among single cells under distinct biological conditions (e.g., different cell types or treatment groups) are calculated in the SHAP embedding space, followed by PERMANOVA multivariate statistical testing to identify genomic regions with statistically significant compartmental alterations between two groups.

Evaluations on both simulated and real scHi-C datasets demonstrate that DeepExDC achieves superior accuracy in detecting diverse patterns of compartmental dynamics. Its identified differential compartments are highly consistent with results generated by state-of-the-art bulk-level methods. Notably, this model operates without restrictive distribution assumptions or predefined differential patterns, endowing it with strong cross-modal transferability. The original framework has been successfully extended to scRNA-seq and scATAC-seq data with robust statistical power, opening new avenues for multi-omics integrated differential analysis of chromatin compartmentalization.

In addition to the four representative methods mentioned above, several other approaches leveraging deep learning strategies merit brief overview. HiCDiffusion, [14] published in BMC Genomics in 2024, combines diffusion models with encoder–decoder neural networks to predict chromatin interaction profiles from DNA sequences, enabling the subsequent inference of compartment architecture. HiCGen, released in Advanced Science in 2025, is a hierarchical prediction model built upon DNA sequences and genomic features, which can delineate genome organization across diverse cell types and spatial scales. Collectively, these studies shape the cutting-edge landscape of deep learning applications in chromatin compartment analysis.

5. Future Research Directions and Development Trends

5.1. Multimodal Data Fusion and Integrated Modeling

Although numerous methods have been developed for chromatin compartment classification based on DNA sequences, histone modifications, or Hi-C contact matrices separately, the systematic fusion of multimodal data remains a core challenge in this research field. Different types of omics data, including Hi-C, ChIP-seq, ATAC-seq, RNA-seq, and DNA methylation profiles, capture distinct dimensions of information regarding chromatin spatial organization. Future methodological trends will focus on constructing deep learning frameworks capable of simultaneously processing and integrating multiple data modalities. Representative strategies involve multi-channel input design and cross-attention mechanisms to achieve semantic alignment across heterogeneous omics datasets. The contrastive learning framework adopted by CDACHIE has preliminarily demonstrated great potential for integrating Hi-C interactome with diverse epigenomic signatures, highlighting a promising developmental trend toward unified multi-omics modeling of chromatin compartmentalization.

5.2. Foundation Models and Cross-Species Generalization

The great success of large language models in natural language processing has sparked widespread interest in genomic foundation models. By treating DNA sequences as a "language of life", the encoding capabilities of large-scale pre-trained Transformer models (e.g., Enformer, DNABERT) can be transferred to 3D genome prediction tasks, which is expected to substantially improve generalization performance across cell types and species. Furthermore, the cross-species transferability between humans and mice demonstrated by SACSANN and ABCRNet indicates that the core sequence determinants governing chromatin compartmentalization are evolutionarily conserved. Developing tools that enable zero-shot or few-shot compartment prediction in non-model species will greatly expand the research boundaries of 3D genome evolution.

5.3. Interpretability and Causal Inference

The transition of compartment prediction models from "black-box" paradigms toward interpretable architectures has emerged as a pivotal trend in this field. The SHAP interpretability framework adopted by DeepExDC offers a powerful tool for unraveling how computational models arrive at specific compartmental assignments. Future advances are expected to shift from correlation-based interpretation to rigorous causal inference. By integrating

paired datasets generated through genetic perturbation, CRISPR genome editing and epigenomic modification experiments, deep learning approaches can be leveraged to dissect the causal effects of specific sequence elements and epigenetic marks on the spatial positioning of chromatin. Progress in this direction will fundamentally deepen mechanistic understanding of 3D genome folding and higher-order chromatin organization.

5.4. Single-Cell and Spatially Resolved Analysis

The rapid advancement of single-cell technologies has provided unprecedented resolution for chromatin compartment analysis. By resolving geometric features of single-cell chromatin structures, MaxComp uncovers widespread cell-to-cell compartment heterogeneity that is masked by conventional population-averaged methods. A key future direction lies in the deep integration of single-cell compartment profiling with spatial transcriptomics and in situ imaging technologies. This integration enables the localization and tracking of spatial dynamics in chromatin compartment states across intact tissue sections, thereby offering unique insights for research on developmental biology, tumor microenvironments, and related biological processes.

5.5. Clinical Translation and Application

Chromatin compartment rearrangement has been widely documented in various cancers and developmental disorders. Most current computational tools are designed for basic research and lack optimization tailored to clinical scenarios [15]. Future efforts should focus on developing robust compartment prediction methods that perform stably on clinical specimens, such as FFPE-fixed tissues and trace biopsy samples. Additionally, quantitative alterations in compartment states need to be systematically evaluated as novel disease biomarkers or predictive indicators of drug response. The scoring strategy of HiC-SCA for assessing compartment assignment quality provides preliminary insights for this translational direction.

6. Conclusion

As a fundamental component of three-dimensional genome analysis, the classification of chromatin A/B compartments witnesses profound paradigm shifts in bioinformatics, evolving from hypothesis-driven to data-driven and further toward model-driven research. This review systematically summarizes the logical evolution of relevant methodologies, ranging from traditional PCA analysis and machine learning to advanced deep learning frameworks. It also comprehensively outlines the design principles, technical characteristics, and applicable scenarios of representative approaches, including SACSANN, HiC-SCA, ABCRNet, MaxComp, ABCNet, CoRNN, TECSAS, and DeepExDC.

These methods exhibit distinct advantages and technical routes. Some aim to infer compartment identity directly from DNA sequences to reduce reliance on Hi-C data. Others leverage epigenomic profiles to achieve stable and robust prediction across diverse cell types. Certain studies reformulate compartment detection as geometric optimization or graph clustering tasks, while specialized interpretable deep learning frameworks are developed to address differential compartment analysis for single-cell datasets.

Current challenges in this field, such as the establishment of unified benchmark criteria, the improvement of model interpretability, the enhancement of cross-cell-type generalization, and the translational advancement toward clinical practice, act as both existing limitations and core driving forces for future methodological innovation. With the continuous maturation of multimodal data fusion strategies, the rapid progress of genomic foundation models, and the further advancement of single-cell and spatially resolved technologies, chromatin compartment classification will evolve from identifying known genomic patterns to uncovering uncharacterized regulatory mechanisms. Such advances will yield in-depth insights into the functional roles of 3D genome organization in both physiological homeostasis and pathological disorders.

References

[1] Lieberman-Aiden E, van Berkum NL, Williams L, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science, 2009, 326(5950): 289-293.

[2] RAMíREZ F, BHARDWAJ V, ARRIGONI L, et al. High-resolution TADs reveal DNA sequences underlying genome organization in flies [J]. Nature communications, 2018, 9(1): 189.

[3] Nattestad M, Schatz M C. Assemblytics: a web analytics tool for the detection of assembly-based variants [J]. Cold Spring Harbor Laboratory, 2016, (19).

[4] Rao S S P, Huntley M H, Durand N C, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping [J]. Cell, 2014, 159(7): 1665-1680.

[5] Prost J, Cameron CJF, Blanchette M. SACSANN: identifying sequence-based determinants of chromosomal compartments. bioRxiv, 2020.

[6] Chan J, Kono H. HiC-SCA: a spectral clustering method for reliable A/B compartment assignment from Hi-C data. bioRxiv, 2025.

[7] Gill R. ABCRNet: predicting membership variability of chromatin compartments. MSc Thesis, University of Guelph, 2025.

[8] Zhan Y, et al. MaxComp: predicting single-cell chromatin compartments from 3D chromosome structures. PLOS Computational Biology, 2025, 21(5): e1013114.

[9] Ashoor H, et al. Sub-Compartment Identifier (SCI): graph embedding and unsupervised learning predict genomic sub-compartments from Hi-C chromatin interaction data. Nature Communications, 2020, 11: 735.

[10] Kirchhof M. ABCNet: A/B compartment prediction from DNA sequence using convolutional neural networks. MSc Thesis, University of Guelph, 2021.

[11] Zheng S, Thakkar N, Harris HL, et al. Predicting A/B compartments from histone modifications using deep learning. iScience, 2024, 27(5): 109570.

[12] Dodero-Rojas E, Contessoto VG, Fehlis Y, et al. Epigenetics is all you need: a transformer to decode chromatin structural compartments from the epigenome. PLOS Computational Biology, 2025, 21(12): e1012326.

[13] Lyu H, Cao P, Long W, et al. DeepExDC interprets genomic compartmentalization changes in single-cell Hi-C data. Briefings in Bioinformatics, 2025, 26(3): bbaf301.

[14] Yoshinaga A, Maruyama O. CDACHIE: chromatin domain annotation by integrating chromatin interaction and

epigenomic data with contrastive learning. Bioinformatics, 2025.

[15] Abdennur N, Mirny L A. Cooler: scalable storage for Hi-C data and other genomically labeled arrays [J]. Bioinformatics, 2019, 36(1): 311-6.