2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) | 979-8-3503-3748-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/BIBM58861.2023.10385614

scHi-CNN: a Computational Method for Statistically Significant Single-cell Hi-C Chromatin Interactions with Nearest Neighbors

Chanaka Bulathsinghalage
Department of Computer Science
North Dakota State University
Fargo, USA
chanaka.cooray@ndsu.edu

Lu Liu
Department of Computer Science
North Dakota State University
Fargo, USA
lu.liu.2@ndsu.edu

***Abstract*—The intricate interplay of regulatory elements, spatial arrangements, and transcription factors shapes the complex chromatin architecture within individual cells, offering valuable insights into cellular diversity and heterogeneity in the realm of chromatin biology. Nevertheless, the analysis of single-cell Hi-C data presents notable challenges due to its sparse nature and limited interaction counts. In this study, we introduce a novel algorithm, scHi-CNN, designed for the detection of statistically significant single-cell Hi-C chromatin interactions. Our method comprises three key steps: imputation of single-cell matrices, normalization, and identification of statistically significant interactions. To assess the robustness and scalability of scHi-CNN across various conditions, we evaluate its performance using three distinct datasets: human cortex cells, mouse embryonic stem cells, and a mouse cell cycle dataset. Moreover, we delve into the biological relevance of the derived significant interactions by examining CTCF binding sites, known promoter-related interactions, and the overlap between different datasets of the same cell type. The results underscore the ability of scHi-CNN to identify more biologically meaningful interactions from single-cell data, facilitating a deeper comprehension of regulatory elements and spatial arrangements within individual cells and across diverse cell types.**

Code and sample data for this paper are available on the GitHub repository at https://github.com/bignetworks2019/scHi-CNN

***Index Terms*—single-cell, Hi-C, chromatin interaction**

I. Introduction

Single-cell chromatin interaction data plays a crucial role in unraveling the intricacies of three-dimensional chromatin structure, capturing cellular heterogeneity, and elucidating genomic variations across diverse cell types. Identifying significant interactions from raw interaction data is imperative for examining regulatory elements, spatial arrangements, transcription factor functions, and other functional elements in individual cells. However, processing single-cell Hi-C data presents several challenges due to its inherent sparseness and limited interaction counts.

Despite the availability of single-cell chromatin interaction datasets to the public, the analysis of significant intra-chromosomal interactions within individual cells is still in its nascent stage. Existing tools primarily focus on imputing and modeling chromatin interactions in single-cell contact matrices, utilizing approaches such as analyzing topologically associating domains, embeddings, and cluster domains [1]–[4]. Furthermore, a computational tool has been developed for identifying frequent inter-chromosomal interactions from single cells using a network-based method [5]. However, none of these tools effectively address the identification of significant intra-chromosomal interactions at the single-cell level. In many cases, researchers resort to employing bulk Hi-C technologies like HiCCUPS [6] and FitHiC [7] to derive significant interactions by aggregating individual cell interactions. Unfortunately, these methods typically yield suboptimal results as they are not tailored to identify significant chromatin interactions specifically within single cells.

Recently, SnapHiC, a random walk algorithm-based method, has been introduced as a pioneering computational approach for identifying significant intra-chromosomal interactions from single-cell Hi-C data [8]. The method has shown promise in enabling the analysis of very high-resolution chromatin interactions (e.g., 10kb) from single-cell Hi-C data. However, the high-resolution nature of these chromatin interactions imposes stringent requirements on the raw single-cell Hi-C data. It is recommended that each single cell possesses a minimum of 150,000 raw chromatin contacts, a threshold that most existing unfiltered single-cell Hi-C data fails to meet. Moreover, SnapHiC treats chromatin interactions in each cell as independent entities, disregarding the local similarities of chromatin interactions between different cells. Notably, leveraging local similarities has proven effective in enhancing the analysis of single-cell Hi-C data [4] and single-cell Hi-C data clustering [9]. Furthermore, the majority of single-cell studies [10]–[16] have been conducted at resolutions of hundreds of kilobases or several megabases. Consequently, there is a need for new computational methods that can accommodate a wider range of single-cell Hi-C data while considering the local similarities of chromatin interactions between different cells, particularly at a comparatively relaxed resolution (e.g., 100kb).

In this study, we propose a novel algorithm for statisti-

cally significant **s**ingle-**c**ell **Hi-C** chromatin interactions with **N**earest **N**eighbors, named **scHi-CNN**. The algorithm comprises three main steps: imputation of single-cell matrices utilizing a k-nearest-neighbor-based approach, normalization, and identification of statistically significant chromatin interactions. To evaluate the performance of our proposed method, we primarily compared it with the SnapHiC algorithm. We utilized three distinct types of single-cell datasets and compared the counts of significant interactions as well as the overlapping interactions between different datasets of the same cell type. Additionally, we assessed the relevance of the derived significant interactions by analyzing CTCF binding sites considering the fact that CTCF plays an important role in three-dimensional genome organization and presumely contributes to the formation of higher-order chromatin structure [17]. To provide a comprehensive comparison, we utilized bulk Hi-C data and contrasted the outcomes obtained from the different methods. Furthermore, we conducted an analysis of chromatin loops generated using varying numbers of cells, focusing on known regulatory elements. The results demonstrated that our proposed algorithm is capable of identifying more biologically meaningful interactions from single-cell data, even when utilizing a smaller number of cells compared to SnapHiC. We firmly believe that our method serves as a valuable tool for identifying significant chromatin interactions in single-cell data, thereby contributing to the analysis of three-dimensional chromatin organization.

II. Method

A. Proposed Algorithm

Our proposed algorithm consists of three key steps: imputation, normalization, and identification of significant chromatin interactions. Our algorithm workflow is visually represented in Fig 1.

1) Imputation of single cell contact matrices: The initial step involved partitioning each chromosome into equal-sized bins for each individual cell. One of our aim is to handle datasets with fewer chromatin interactions, so we've chosen to broaden the bin size resolution of SnapHiC from high resolution bins (like 10Kb, 25Kb) to 100Kb. We then assigned chromatin interactions to specific bin pairs and tallied these interactions to generate contact matrices. For contact matrices that contained empty pixels (i.e., zero contact count), we implemented a strategy to impute these empty pixels. Specifically, we extracted a surrounding region measuring *(2d+1)x(2d+1)* (e.g *d=5* bin pair differences in each direction from the empty pixel) to identify the closest neighbors. To perform imputation, we only considered pixels that had at least one chromatin interaction within their surrounding region.

Subsequently, we retrieved the surrounding matrices corresponding to the same position in the other cells for the same chromosome. From these matrices, we selected the top k (e.g k=4, which is also used in [1]) neighbors based on the Pearson correlation coefficient. The mean of these top k closest neighbors was then used to impute the empty pixel. Note that after the imputation, the empty pixel can still be zero if the same entries are zeros for all top k neighbors.

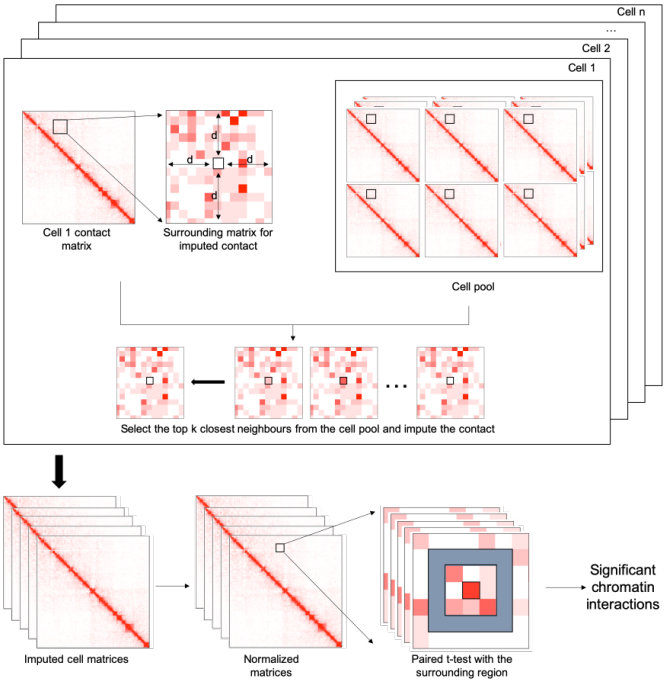

Fig. 1. Workflow of the Method - 1. Single-cell contact matrix imputation, 2. Normalization process, 3. Identification of significant chromatin interactions

To maintain the integrity of the analysis, we imposed a maximum distance threshold (e.g. 1 million base pairs) for the imputation of interactions. This ensures that the imputed values are derived from nearby genomic regions that are more likely to exhibit chromatin interactions. Also, given the symmetry of a Hi-C matrix, our procedure involved only the imputation of the upper half matrix within the specified distance.

2) Normalization: To standardize our contact matrices, we employed a normalization approach that involved grouping interactions with the same genomic distance, effectively normalizing them diagonally with the same parameters used in SnapHiC [8] for a fair comparison. For each diagonal segment within a contact matrix, we started by filtering the top 1% of the interactions with the highest contact values. Subsequently, we computed the mean and standard deviation using the remaining values and calculated corresponding z-scores. Diagonals with a standard deviation lower than 10^{-6} were disregarded, and those segments were filled with zeros to account for their negligible variability.

3) Identification of significant chromatin interactions: To identify significant chromatin interactions , we implemented similar criteria used in SnapHiC to determine if a interaction bin pair qualified as a peak compared to its surrounding region. For an interaction pair to be considered, its mean normalized contact counts across all cells needed to exceed zero. Additionally, we require that at least 10% of single cells exhibited a normalized contact count greater than 1.96

(corresponding to a $pvalue \leq 0.05$). For interactions that meet these criteria, we conducted a paired t-test with the local neighborhood to assess significance. The local neighborhood was defined as the surrounding regions within a 2-bin genomic distance, excluding the closest neighbors (i.e., bin pairs within a 1-bin genomic difference). Using the mean of the local neighborhood values, we performed the paired t-test and obtained t-statistics and p-values. Subsequently, we grouped the p-values based on genomic distance and converted them into false discovery rates (FDRs) using the Benjamini-Hochberg procedure. Finally, we identified the significant chromatin interactions based on a t-statistic greater than 3 and an FDR value less than 0.1.

B. Processing Single-Cell Hi-C data

In this study, we utilized several publicly available single-cell Hi-C datasets. Firstly, for the cell cycle dataset [10], we obtained contact matrices for single cells categorized into four distinct cell cycle phases. The labels G1 phase, Early-S phase, Mid-S phase, and G2 phase correspond to the datasets 1CDX1, 1CDX2, 1CDX3, and 1CDX4 respectively. Each phase included interaction data for a total of 390 individual cells. Secondly, we acquired contact matrices for Mouse ES cells [18] comprising a total of 475 cells. Lastly, we obtained contact matrices for human frontal cortex single cells [18] that comprise a total of 4,238 cells. To process the single-cell Hi-C data, we applied both the proposed algorithm and the SnapHiC algorithm, allowing for a comparison of the results obtained from each method.

C. Processing Bulk Hi-C data

In the study, we obtained the Fastq files for the bulk Hi-C data [19] corresponding to the cell cycle dataset. These files were then processed using HiC-Pro to generate contact matrices [20]. For the bulk Hi-C data related to Mouse ES cells [18], we directly downloaded the contact matrices from the NCBI database. To identify significant chromatin interactions within these contact matrices, we applied the HiCCUPS [6] and FitHiC2 [7] algorithms. In order to compare these findings with the single-cell Hi-C data, we focused on the common interactions identified by both HiCCUPS and FitHiC2 algorithms.

D. Processing CTCF ChIP-seq Data

The Mouse ES cells CTCF ChIP-seq narrow peak data were obtained from the ENCODE project (ENCSR362VNF) [21]. Similarly, for Homo sapiens neural cells derived from H1, the CTCF ChIP-seq data were downloaded from ENCODE (ENCSR822CEA). To analyze the single-cell Hi-C datasets, we performed a counting of CTCF-enriched interactions. An interaction was classified as CTCF-enriched if both ends of the interaction overlapped with at least one CTCF binding site. This criterion allowed us to identify and examine interactions that exhibited a potential association with CTCF binding events.

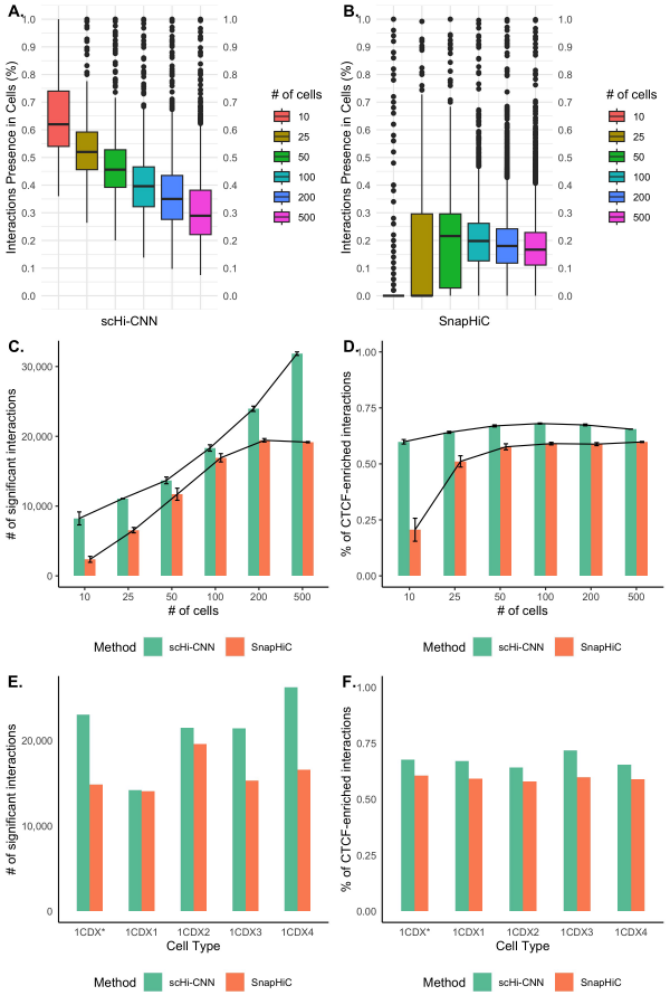

Fig. 2. A. Distribution of the percentages of the presence of raw interactions corresponding to the identified significant interactions across cells in prefrontal cortex for scHi-CNN (e.g., 0.5 means 50% of the cells contain the identified significant interaction) B. Same as 'A' for the SnapHiC method C. Significant interactions derived using scHi-CNN and SnapHiC for cells in prefrontal cortex. D. Percentage of CTCF enriched interactions identified using the two methods for cells in prefrontal cortex. In A,B,C, and D five random samples for each number of cells were gathered and represented in the figure with the error bars. E. Significant interactions derived using two methods for cell cycle data organized in each cell cycle. F. Percentage of CTCF enriched interactions identified using the two methods for cell cycle data.

E. Processing promoter related interactions

In this study, we utilized a previously reported set of promoter-related interactions, including promoter-promoter and promoter-other interactions, as a reference dataset [22]. These interactions were derived from a study conducted on human cortex cells. To evaluate the performance of our proposed methodology, we compared our results with those obtained from SnapHiC with varying numbers of cells. We then examined the overlap between these interactions and the reference promoter-related interactions. This analysis allowed us to assess the accuracy and effectiveness of our methodology in capturing relevant chromatin interactions within the context of promoter activity in human cortex cells.

III. Results

A. Quantity of significant chromatin interactions

For the analysis of human cortex cells, we used both methodologies across varying cell numbers, namely 10, 25, 50, 100, 200, and 500 cells. To ensure unbiased and representative results, we performed multiple random selections of cell numbers, as depicted by the error bars in Figure 2C. The outcomes consistently demonstrate that scHi-CNN identifies a significantly higher number of chromatin loops, even when applied to a small cell population. In contrast, SnapHiC's performance appears to be less effective, particularly in detecting interactions among smaller cell groups. To extend our evaluation, we applied both methodologies to the whole and each cell phase in the cell cycle dataset. The performance remains consistent across the cell phases, as illustrated in Figure 2E. Also scHi-CNN is capable of identifying the increasing trend of significant chromatin interactions in cells at varying stages of the cell cycle, a phenomenon attributed to DNA replication during the S phase. In contrast, SnapHiC is unable to capture these inherent biological states of the cell. Furthermore, we quantified the raw interactions corresponding to the identified significant interactions (Figure 2A and B). Notably, when analyzing smaller cell groups (around 10 cells), scHi-CNN identifies interactions that were present in approximately 60% of the cells, whereas the interactions identified by the SnapHiC method are present in a much smaller fraction of cells. This showcases the superiority of scHi-CNN in identifying frequently occurring chromatin interactions among cells, thereby highlighting its potential to derive more relevant chromatin loops.

B. CTCF enriched interactions

CTCF plays an important role in three-dimensional genome organization and presumely contributes to the formation of higher-order chromatin structure [17]. We assessed the CTCF enrichment of the significant interactions obtained from scHi-CNN and SnapHiC by leveraging previously collected CTCF methylation data (Figure 2D,F). Our analysis reveals that the percentage of CTCF-enriched interactions derived from scHi-CNN remains consistent across different cell quantities, whereas SnapHiC struggles to generate CTCF-enriched interactions, especially when dealing with smaller cell populations. SnapHiC requires a minimum of 50-100 cells to produce 50% of the CTCF-enriched interactions. In contrast, scHi-CNN consistently identifies more than 60% of CTCF-enriched interactions in both human cortex cell and cell cycle datasets, regardless of the number of cells used. These findings suggest that the results obtained using scHi-CNN encompass a higher proportion of biologically meaningful data, indicating an improvement over existing methodologies in terms of data quality and relevance.

C. Common interactions between different datasets from the same cell type

To investigate the overlap of interactions between different datasets, we analyzed the cell cycle dataset and mouse embryonic stem cell (mESC) dataset, which used the same cell type. We specifically examined the common interactions between each phase of the cell cycle and mESC cells for both scHi-CNN and SnapHiC (Figure 3). Additionally, we determined the common interactions across each cell phase within the cell cycle dataset (Figure 4). Interestingly, scHi-CNN consistently identifies a significantly higher percentage of common interactions in both cases. In addition, the Figure 3 illustrates that scHi-CNN outperforms SnapHiC in terms of stability, as evidenced by a lower maximum variation in the common percentage (6.36% for scHi-CNN compared to 7.86% for SnapHiC). This observation suggests that scHi-CNN excels in deriving meaningful interactions by effectively identifying a greater number of common interactions within the same cell type.

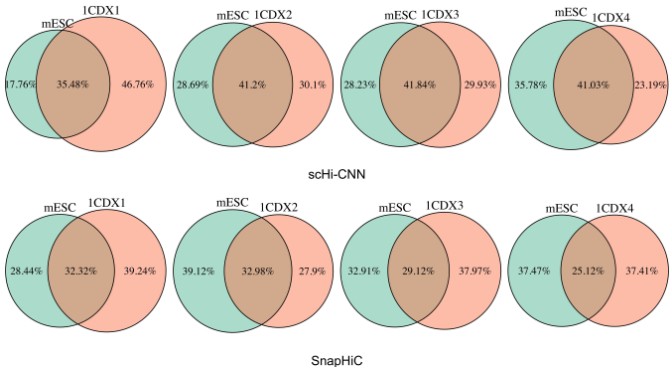

Fig. 3. Common interactions percentages between the cell cycle and mESC datasets using scHi-CNN and SnapHiC

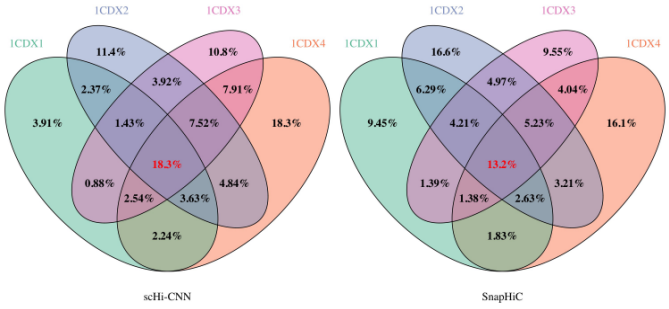

Fig. 4. Common interactions percentages among cell cycle phases using scHi-CNN and SnapHiC

D. Identified promoter centered interactions

To gain further insights into the identified interactions, we conducted an evaluation using Layer 2/3 (L2/3) type cells from human cortex cells, considering different quantities of cells. In order to facilitate more comparison, we also employed SnapHiC at 10kb resolution with 100 L2/3 cells. We specifically focused on four known promoters and genes associated with cortex and neural cells, as highlighted in previous studies [22]–[24], to assess the identified chromatin

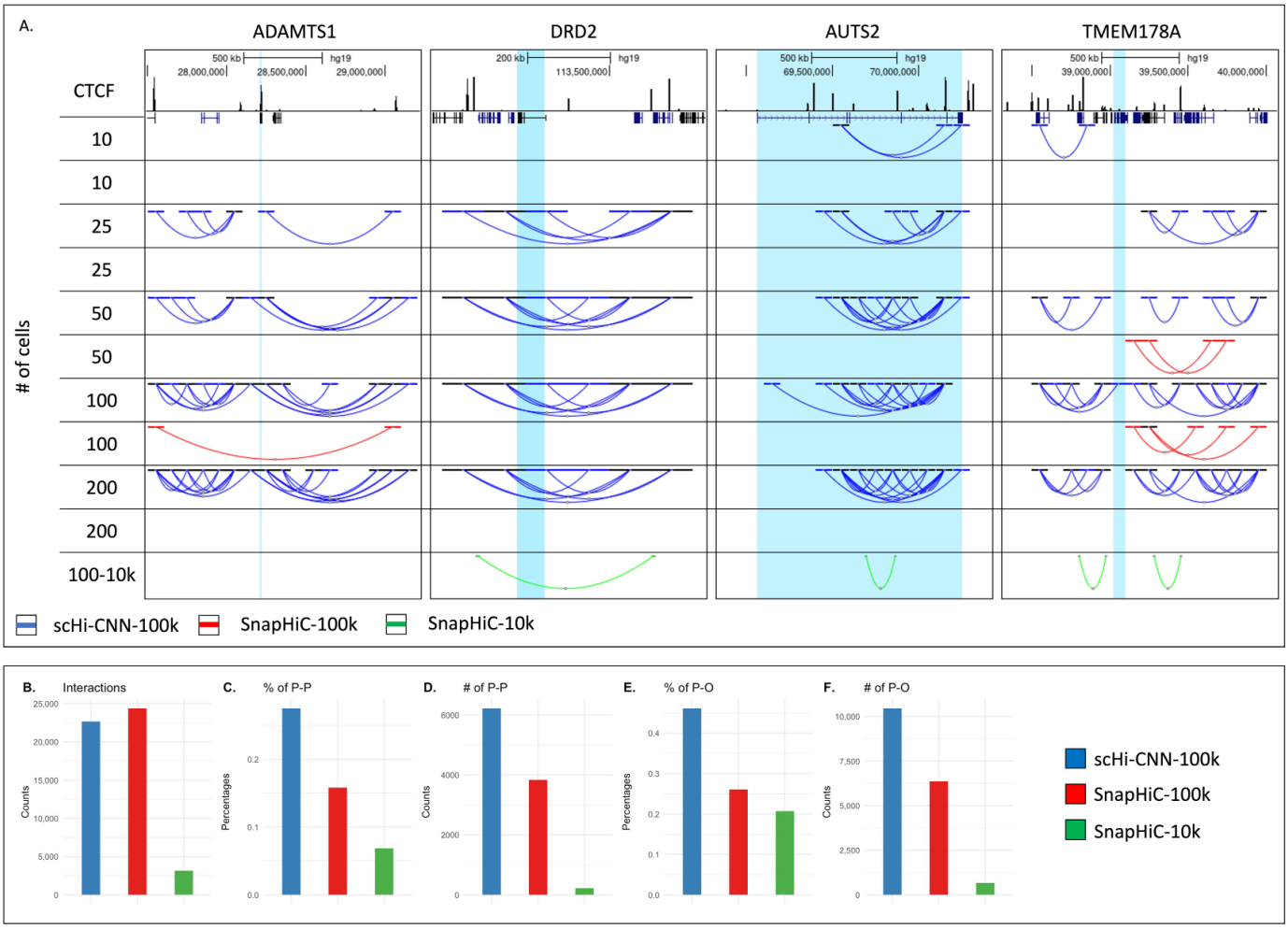

Fig. 5. Identified significant interactions in human cortex cell lines related to known Promoter-centered interactions using scHi-CNN and SnapHiC. A. Identified significant interactions for each cell quantity using scHi-CNN and SnapHiC within the marked areas associated with the four known promoters. B. Number of significant interactions derived using scHi-CNN and SnapHiC. C and E. Percentage of overlap with known promoter-promoter interactions and promoter-other interactions. D and F. Overlapping interaction count with known promoter-promoter interactions and promoter-other interactions.

interactions. Figure 5A showcases the identified interactions for each cell quantity using scHi-CNN, while highlighting the promoters of interest. Remarkably, scHi-CNN successfully identifies these promoter-related interactions even with a very low cell count, whereas SnapHiC fails to detect most of these interactions even with a higher cell count at 100kb resolution. Although SnapHiC manages to identify a few promoter-related interactions at 10kb resolution, its performance fells short compared to scHi-CNN. Furthermore, Figure 5C,D,E,F illustrates the overlapping promoter-centered interactions identified using 100 cortex single cells, in comparison with the promoter-centered interactions reported in a previous study [22]. Though scHi-CNN identifies less number of significant interactions (Figure 5B) than SnapHiC, our method, scHi-CNN, reports a significantly higher percentage of promoter-centered interactions compared to SnapHiC (Figure 5C,D,E,F). These findings further highlight the superior performance of scHi-CNN in identifying a greater proportion of biologically meaningful interactions.

IV. Conclusion

In conclusion, this study presents a novel and robust methodology for identifying significant intra-chromosomal chromatin loops from single-cell Hi-C data, addressing the limitations of existing tools and expanding our understanding of chromatin architecture in individual cells. Our method consists of three primary steps: 1) imputing contact matrices using a K-nearest-neighbour-based approach, 2) normalization, and 3) identifying significant chromatin interactions using a statistical test. We evaluated the performance of our proposed approach using three distinct datasets, including human cortex cells, mouse embryonic stem (ES) cells, and a mouse cell cycle dataset, with varying numbers of cells to assess the robustness and scalability of our method across different conditions.

To validate the biological relevance of the interactions identified by our approach, we utilized several criteria, including CTCF binding sites, analysis of known promoter-related interactions, and quantification of common interactions between different datasets of the same cell type. Our method shows a

greater ability to generate a significantly higher number of biologically meaningful interactions compared to SnapHiC. The capabilities were demonstrated through a higher percentage of CTCF-enriched interactions, greater overlap with known promoter-centered interactions, and increased common interactions between the same cell types, thus highlighting the potential of our method in deciphering complex regulatory networks in single cells.

Future research could focus on refining and optimizing the methodology to further enhance its performance, sensitivity, and generalizability across diverse cell types and conditions. Additionally, integrating our method with other single-cell genomics data modalities, such as single-cell RNA-seq, ATAC-seq, or ChIP-seq, could provide a more comprehensive view of the molecular mechanisms associating with chromatin architecture and gene regulation in single cells. This multi-modal integration would enable researchers to better understand the complex interplay between chromatin structure and function, ultimately leading to novel therapeutic strategies for various diseases, including cancer and developmental disorders, which are often characterized by aberrant chromatin organization and gene expression patterns.

ACKNOWLEDGMENT

The work is supported by the National Science Foundation under NSF EPSCoR Track-1 Cooperative Agreement OIA #1946202 and used advanced cyberinfrastructure resources provided by the University of North Dakota Computational Research Center and the Center for Computationally Assisted Science and Technology (CCAST) at North Dakota State University.

REFERENCES

[1] J. Zhou, J. Ma, Y. Chen, C. Cheng, B. Bao, J. Peng, T. J. Sejnowski, J. R. Dixon, and J. R. Ecker, "Robust single-cell hi-c clustering by convolution-and random-walk–based imputation," *Proceedings of the National Academy of Sciences*, vol. 116, no. 28, pp. 14 011–14 018, 2019.

[2] X. Li, F. Feng, H. Pu, W. Y. Leung, and J. Liu, "schictools: a computational toolbox for analyzing single-cell hi-c data," *PLoS computational biology*, vol. 17, no. 5, p. e1008978, 2021.

[3] Q. Xie, C. Han, V. Jin, and S. Lin, "Hicimpute: A bayesian hierarchical model for identifying structural zeros and enhancing single cell hi-c data," *PLoS computational biology*, vol. 18, no. 6, p. e1010129, 2022.

[4] R. Zhang, T. Zhou, and J. Ma, "Multiscale and integrative single-cell hi-c analysis with higashi," *Nature biotechnology*, vol. 40, no. 2, pp. 254–261, 2022.

[5] C. Bulathsinghalage and L. Liu, "Network-based method for regions with statistically frequent interchromosomal interactions at single-cell resolution," *BMC bioinformatics*, vol. 21, no. 14, pp. 1–15, 2020.

[6] S. S. Rao, M. H. Huntley, N. C. Durand, E. K. Stamenova, I. D. Bochkov, J. T. Robinson, A. L. Sanborn, I. Machol, A. D. Omer, E. S. Lander *et al.*, "A 3d map of the human genome at kilobase resolution reveals principles of chromatin looping," *Cell*, vol. 159, no. 7, pp. 1665–1680, 2014.

[7] F. Ay, T. L. Bailey, and W. S. Noble, "Statistical confidence estimation for hi-c data reveals regulatory chromatin contacts," *Genome research*, vol. 24, no. 6, pp. 999–1011, 2014.

[8] M. Yu, A. Abnousi, Y. Zhang, G. Li, L. Lee, Z. Chen, R. Fang, T. M. Lagler, Y. Yang, J. Wen *et al.*, "Snaphic: a computational pipeline to identify chromatin loops from single-cell hi-c data," *Nature methods*, vol. 18, no. 9, pp. 1056–1059, 2021.

[9] J. Wolff, R. Backofen, and B. Grüning, "Robust and efficient single-cell hi-c clustering with approximate k-nearest neighbor graphs," *Bioinformatics*, vol. 37, no. 22, pp. 4006–4013, 2021.

[10] T. Nagano, Y. Lubling, C. Várnai, C. Dudley, W. Leung, Y. Baran, N. Mendelson Cohen, S. Wingett, P. Fraser, and A. Tanay, "Cell-cycle dynamics of chromosomal organization at single-cell resolution," *Nature*, vol. 547, no. 7661, pp. 61–67, 2017.

[11] V. Ramani, X. Deng, R. Qiu, K. L. Gunderson, F. J. Steemers, C. M. Disteche, W. S. Noble, Z. Duan, and J. Shendure, "Massively multiplex single-cell hi-c," *Nature methods*, vol. 14, no. 3, pp. 263–266, 2017.

[12] I. M. Flyamer, J. Gassler, M. Imakaev, H. B. Brandão, S. V. Ulianov, N. Abdennur, S. V. Razin, L. A. Mirny, and K. Tachibana-Konwalski, "Single-nucleus hi-c reveals unique chromatin reorganization at oocyte-to-zygote transition," *Nature*, vol. 544, no. 7648, pp. 110–114, 2017.

[13] T. J. Stevens, D. Lando, S. Basu, L. P. Atkinson, Y. Cao, S. F. Lee, M. Leeb, K. J. Wohlfahrt, W. Boucher, A. O'Shaughnessy-Kirwan *et al.*, "3d structures of individual mammalian genomes studied by single-cell hi-c," *Nature*, vol. 544, no. 7648, pp. 59–64, 2017.

[14] H.-J. Kim, G. G. Yardımcı, G. Bonora, V. Ramani, J. Liu, R. Qiu, C. Lee, J. Hesson, C. B. Ware, J. Shendure *et al.*, "Capturing cell type-specific chromatin compartment patterns by applying topic modeling to single-cell hi-c data," *PLoS computational biology*, vol. 16, no. 9, p. e1008173, 2020.

[15] T. Nagano, Y. Lubling, T. J. Stevens, S. Schoenfelder, E. Yaffe, W. Dean, E. D. Laue, A. Tanay, and P. Fraser, "Single-cell hi-c reveals cell-to-cell variability in chromosome structure," *Nature*, vol. 502, no. 7469, pp. 59–64, 2013.

[16] R. M. Mulqueen, D. Pokholok, B. L. O'Connell, C. A. Thornton, F. Zhang, B. J. O'Roak, J. Link, G. G. Yardımcı, R. C. Sears, F. J. Steemers *et al.*, "High-content single-cell combinatorial indexing," *Nature biotechnology*, vol. 39, no. 12, pp. 1574–1580, 2021.

[17] G. Fudenberg, M. Imakaev, C. Lu, A. Goloborodko, N. Abdennur, and L. A. Mirny, "Formation of chromosomal domains by loop extrusion," *Cell reports*, vol. 15, no. 9, pp. 2038–2049, 2016.

[18] D.-S. Lee, C. Luo, J. Zhou, S. Chandran, A. Rivkin, A. Bartlett, J. R. Nery, C. Fitzpatrick, C. O'Connor, J. R. Dixon *et al.*, "Simultaneous profiling of 3d genome structure and dna methylation in single human cells," *Nature methods*, vol. 16, no. 10, pp. 999–1006, 2019.

[19] B. Bonev, N. M. Cohen, Q. Szabo, L. Fritsch, G. L. Papadopoulos, Y. Lubling, X. Xu, X. Lv, J.-P. Hugnot, A. Tanay *et al.*, "Multiscale 3d genome rewiring during mouse neural development," *Cell*, vol. 171, no. 3, pp. 557–572, 2017.

[20] N. Servant, N. Varoquaux, B. R. Lajoie, E. Viara, C.-J. Chen, J.-P. Vert, E. Heard, J. Dekker, and E. Barillot, "Hic-pro: an optimized and flexible pipeline for hi-c data processing," *Genome biology*, vol. 16, no. 1, pp. 1–11, 2015.

[21] E. P. Consortium *et al.*, "An integrated encyclopedia of dna elements in the human genome," *Nature*, vol. 489, no. 7414, p. 57, 2012.

[22] I. Jung, A. Schmitt, Y. Diao, A. J. Lee, T. Liu, D. Yang, C. Tan, J. Eom, M. Chan, S. Chee *et al.*, "A compendium of promoter-centered long-range chromatin interactions in the human genome," *Nature genetics*, vol. 51, no. 10, pp. 1442–1449, 2019.

[23] M. Song, X. Yang, X. Ren, L. Maliskova, B. Li, I. R. Jones, C. Wang, F. Jacob, K. Wu, M. Traglia *et al.*, "Mapping cis-regulatory chromatin contacts in neural cells links neuropsychiatric disorder risk variants to target genes," *Nature genetics*, vol. 51, no. 8, pp. 1252–1262, 2019.

[24] X. Luo, Y. Liu, D. Dang, T. Hu, Y. Hou, X. Meng, F. Zhang, T. Li, C. Wang, M. Li *et al.*, "3d genome of macaque fetal brain reveals evolutionary innovations during primate corticogenesis," *Cell*, vol. 184, no. 3, pp. 723–740, 2021.