

# CS 6000 - Assignment - Journal #3

Carlos Eugenio Lopes Pires Xavier Torres

September 11, 2023

## 1 Define research area and search for papers in the area

### 1.1 Research area

My research area in the Bioinformatics field is 3D genome structure reconstruction computational method development focused on single cells.

### 1.2 Search for survey papers

For this search, I used the terms “**single-cell 3D genome structure reconstruction algorithms survey OR review**”, filtered by date inside the last 7 years, to obtain a specific list of related survey/review papers in my area. I also included a few papers that my advisor suggested me to review. I selected the papers that I deemed more interesting and related to my area and added to a Google Scholar library label I created for this assignment. Then I exported the library in BiBTeX to add to my paper references. This is the list I selected:

- [1] An overview of methods for reconstructing 3-D chromosome and genome structures from Hi-C data.
- [2] Computational methods for predicting 3D genomic organization from high-resolution chromosome conformation capture data.
- [3] The 3D genome structure of single cells.
- [4] Techniques for and challenges in reconstructing 3D genome structures from 2D chromosome conformation capture data.
- [5] Comparison of computational methods for 3D genome analysis at single-cell Hi-C level.
- [6] Perspectives for the reconstruction of 3D chromatin conformation using single cell Hi-C data.
- [7] Mapping nucleosome and chromatin architectures: A survey of computational methods.
- [8] Manifold learning-based methods for analyzing single-cell RNA-sequencing data.
- [9] Single-cell Hi-C data analysis: safety in numbers.
- [10] Advances in technologies for 3D genomics research.
- [11] The two waves in single-cell 3D genomics.
- [12] Recent advances in single-cell sequencing technologies.

### 1.3 Git availability

This journal is available in a public Git repository I created in my GitHub account via the integration with Overleaf:

- <https://github.com/cetorres/CS-6000---Assignment---Journal-3>

## 2 Critically/creatively read survey papers

I selected these 5 papers to critically/creatively read for this assignment. I added the information about number of citations and the year they were released:

1. [1] An overview of methods for reconstructing 3-D chromosome and genome structures from Hi-C data. (cited by 103 since 2019)
2. [2] Computational methods for predicting 3D genomic organization from high-resolution chromosome conformation capture data. (cited by 21 since 2020)
3. [5] Comparison of computational methods for 3D genome analysis at single-cell Hi-C level. (cited by 12 since 2020)
4. [10] Advances in technologies for 3D genomics research. (cited by 12 since 2020)
5. [11] The two waves in single-cell 3D genomics. (cited by 8 since 2022)

Below I provide summary and notes about the selected papers:

1. **Notes:** This paper is the most cited of the selected that is at least 3 years old. I believe it is so because it has some of the most important characteristics that we saw in class about great survey paper: it is an overview, introduction, kind of survey paper, which invites more beginners that want to start in the field; it possesses a great organization of the content, very easy to understand and comprehensive; it separates the different methods in distinguished classes: distance based, contact based, and probability based; presents a comprehensive comparison table of the methods; it was released in a period that there were not so many survey papers like this about the field; and the quality of the work overall is superior, not just presenting the techniques, but discussing them, systematically comparing them, and proposing different approaches on the fields resources to better understanding the techniques.

**Summary:** In this paper, the authors provide a comprehensive review of computational methodologies employed for reconstructing 3D chromosome and genome structures using Hi-C data. The authors systematically discuss the key techniques and algorithms used in this field, emphasizing their underlying principles and applications. They highlight the diversity of approaches, including probabilistic modeling, dimensionality reduction, and graph-based representations, each offering unique insights into the spatial organization of chromosomes within the cell nucleus. Furthermore, the paper underscores the importance of considering biological constraints and integrating multi-omics data to refine and validate reconstructed structures. The authors delve into the strengths and limitations of various computational tools and methods used in 3D genome reconstruction, aiding researchers in selecting the most suitable techniques for their specific research objectives. They emphasize the importance of benchmarking and validation, which are critical for ensuring the accuracy and reliability of reconstructed 3D structures. Additionally, the paper underscores the ongoing challenges in the field, such as dealing with noisy data and the need for improved computational efficiency. Overall, this paper provides an invaluable resource for the bioinformatics community, offering a comprehensive overview of the state-of-the-art methods for reconstructing 3D chromosome and genome structures from Hi-C data and guiding future directions in this rapidly evolving field. In summary, this review serves as a valuable resource for researchers in the field, offering a comprehensive understanding of the evolving landscape of 3D genome reconstruction techniques and their potential contributions to advancing our knowledge of genome organization and function.

2. **Notes:** This paper, although is 3 years old, does not have as much citations of the previous on item 1. I believe because it lacks some of the key aspects that make a great survey paper were not applied. Although it has a good quality and present some important computational methods used for reconstructing 3D chromosome structures using Hi-C data, the organization is not the best, it is a more confusing than helpful. Also the title is not inviting for a beginner in the field, because it does not appeal as an introductory work, where the reader would be more gently initiated to the field. Instead it just states what it is going to present very directly, like a research paper.

**Summary:** In this paper, the authors provide a comprehensive exploration of computational techniques designed for the prediction of 3D genomic organization using high-resolution chromosome conformation capture (Hi-C) data. The paper systematically reviews various computational methodologies, focusing on their underlying algorithms and data preprocessing steps. It highlights the significance of accurately capturing

chromatin interactions and structural features from Hi-C data, which is crucial for understanding the spatial organization of the genome within the cell nucleus. The authors discuss the strengths and limitations of different computational tools, shedding light on their performance in accurately deciphering the intricate spatial arrangements of the genome. Moreover, the paper emphasizes the importance of integrating multi-omics data and considering biological context to enhance the accuracy of 3D genomic predictions. It underscores how these computational methods play a pivotal role in advancing our understanding of gene regulation, genome function, and their implications in various biological processes. This paper serves as a valuable resource for bioinformaticians and researchers in genomics, providing a comprehensive overview of state-of-the-art computational tools and methodologies for predicting 3D genomic organization, ultimately contributing to advancements in our understanding of the genome's spatial architecture.

3. **Notes:** This paper (as the one on item 4) has the smallest number of citations of the selected ones in the last 3 years. It falls in the same category of the previews one on item 2. Although it has a better title referring to a comparison of computational methods, which is a nice beginner invitation, it lacks on the organization and ease of understanding.

**Summary:** In this paper, the authors conduct a thorough comparative analysis of various computational methods aimed at elucidating 3D genome structures from single-cell Hi-C data. They systematically evaluate the performance of these methods, considering their abilities to accurately capture chromatin interactions and structural features within individual cells while also taking into account factors like computational efficiency and usability. Through this rigorous benchmarking process, they provide valuable insights into the strengths and limitations of each approach, helping researchers select the most appropriate tools for their specific research questions and datasets. The paper underscores the importance of addressing the unique challenges posed by single-cell Hi-C data, such as increased sparsity and noise, and discusses how different computational methods have tackled these challenges. Additionally, it highlights the potential impact of single-cell 3D genome analysis on our understanding of cellular heterogeneity, gene regulation, and disease mechanisms. Overall, this comprehensive comparative study in the field of bioinformatics contributes significantly to advancing the state of 3D genome analysis at the single-cell level, facilitating more accurate and insightful investigations into the intricate spatial organization of the genome within individual cells.

4. **Notes:** Again, this paper is similar to the paper on item 3. Although its content is introductory and discuss a review and the advancements on the field of 3D genomics research, I believe it did not get too many citations because of the lack of organization structure. The classification and selected examples were not so clear and the overall quality was not as good as the one on item 1.

**Summary:** In this paper, the authors provide a comprehensive overview of recent technological advancements that have significantly contributed to the field of 3D genomics research. They discuss a range of innovative techniques and tools that have emerged in recent years, revolutionizing our ability to investigate the three-dimensional organization of the genome. These advancements encompass improvements in chromatin conformation capture methods, such as Hi-C and its variants, which have enabled researchers to probe genome architecture at unprecedented resolutions. Additionally, the paper highlights the role of cutting-edge imaging and microscopy techniques in visualizing genome structures within the cellular nucleus, shedding light on the spatial organization of chromosomes and their functional implications. Furthermore, the authors emphasize how these technological breakthroughs have advanced our understanding of fundamental biological processes, including gene regulation, genome folding, and epigenetic modifications. They discuss the potential applications of 3D genomics research in diverse areas, such as developmental biology, disease mechanisms, and personalized medicine. This paper serves as a valuable resource for researchers in the field, offering insights into the current state of the art in 3D genomics technologies and their transformative potential for unlocking new insights into genome organization and function. Overall, the paper underscores the pivotal role of technological innovations in driving progress in genomics research and highlights the exciting avenues that lie ahead in this rapidly evolving field.

5. **Notes:** This paper has a nice introductory perspective on the evolution of single-cell 3D genomics research field. It also has a nice organization and classification of its development, and uses a great story-telling writing technique that separate the history of the field in two waves of evolution. That aspect helps the reader understand better the subjects and also maintains the user interested during the whole reading of the paper. It has only a few citations yet because it is fairly newly released, only 1 year old. But I believe it will have more citations with time.

**Summary:** In this review paper, the authors present a compelling perspective on the evolution of single-cell 3D genomics research, categorizing its development into two distinct waves. The authors first describe the initial wave, which primarily focused on examining chromatin accessibility and gene expression at the single-cell level. They highlight that this phase provided valuable insights into cellular heterogeneity but was limited in providing detailed spatial information about the genome’s organization within individual cells. The second wave, as outlined in the paper, represents a significant advancement in the field. It is characterized by the emergence of high-resolution single-cell chromatin conformation capture techniques, such as single-cell Hi-C. These techniques have revolutionized our ability to investigate the 3D genome architecture at unprecedented resolutions within individual cells. The authors emphasize the transformative impact of this second wave, which allows researchers to gain deeper insights into the principles governing genome folding, spatial interactions, and their functional consequences. This paper provides a valuable historical framework for understanding the progression of single-cell 3D genomics research, highlighting the importance of high-resolution techniques in unraveling the complexities of genome organization at the single-cell level and its potential to address fundamental questions in biology and disease.

### 3 Build a topic map of where there are gaps in my research area

The single-cell 3D genome structure reconstruction area has recently grown considerably in the last four years. Also known as single-cell Hi-C (scHi-C), the technologies have enabled the analysis of 3D genome organization at an unprecedented resolution with a readout of chromatin interactions in individual cells. These new sequencing-based methods allow researchers to unveil the dynamics of chromatin conformation at a single-cell level and establish connections between structure and function obscured by population-level genomic mapping methods.

Survey papers that introduce and compare population-level genomic mapping methods, or 3D full genome reconstruction computation methods, are abundant in the field. Because there are many methods already discussed, analyzed, tested, and compared. But for single-cell Hi-C, there are still quite few of them because this sub area is still growing in popularity and interest. And it follows the advancements on the biological field of the actual biochemistry sequencing processes.

The fact that Hi-C data is noisy, coupled with other technical factors, makes it extremely difficult to determine the various unique 3D structures of cells used in Hi-C experiments. Due to the drawbacks involved in using multi-cell Hi-C data, studying single-cell Hi-C data has become increasingly relevant. In particular, it does not require designing an algorithm to satisfy the variability of each cell used in the Hi-C experiment. As expected, single-cell Hi-C datasets are sparser than multi-cell Hi-C datasets. Hence, conventional distance- and restraint-based methods are not suitable for 3D structure reconstruction based on these data.

In conclusion, the are of single-cell Hi-C (scHi-C) 3D genome structure reconstruction is still growing and need more knowledge spreading papers, such as survey papers, to bring more people to it and facilitate the developments of new and improved computational methods. That is why I believe it is a gap worth to explore and write survey papers about.

### 4 Learning process of survey papers search and review

The learning process for this assignment was quite interesting and, in my opinion, will help a lot in a real world research to write a survey paper in my area. The steps I followed to build this assignment, together with the subjects I saw in class about survey papers, how to identify and write good ones, helped me understand how to start to write such kind of paper.

The definition of the research area, then the search for papers, browsing and scanning interesting ones related to the area, then critically/creatively reading some selected ones and analysing why they were more cited or not helped me get the “lay of the land” of survey papers. I believe after this assignment I am more prepared to start writing my first survey paper in my research area and have a successful publication.

## References

- [1] O. Oluwadare, M. Highsmith, and J. Cheng, “An overview of methods for reconstructing 3-d chromosome and genome structures from hi-c data,” *Biological procedures online*, vol. 21, no. 1, pp. 1–20, 2019.

- [2] K. MacKay and A. Kusalik, “Computational methods for predicting 3d genomic organization from high-resolution chromosome conformation capture data,” *Briefings in functional genomics*, vol. 19, no. 4, pp. 292–308, 2020.
- [3] T. Zhou, R. Zhang, and J. Ma, “The 3d genome structure of single cells,” *Annual review of biomedical data science*, vol. 4, pp. 21–41, 2021.
- [4] Z. Li, S. Portillo-Ledesma, and T. Schlick, “Techniques for and challenges in reconstructing 3d genome structures from 2d chromosome conformation capture data,” *Current Opinion in Cell Biology*, vol. 83, p. 102209, 2023.
- [5] X. Li, Z. An, and Z. Zhang, “Comparison of computational methods for 3d genome analysis at single-cell hi-c level,” *Methods*, vol. 181, pp. 52–61, 2020.
- [6] P. I. Kos, A. A. Galitsyna, S. V. Ulianov, M. S. Gelfand, S. V. Razin, and A. V. Chertovich, “Perspectives for the reconstruction of 3d chromatin conformation using single cell hi-c data,” *PLoS Computational Biology*, vol. 17, no. 11, p. e1009546, 2021.
- [7] K. Fang, J. Wang, L. Liu, and V. X. Jin, “Mapping nucleosome and chromatin architectures: A survey of computational methods,” *Computational and Structural Biotechnology Journal*, 2022.
- [8] K. R. Moon, J. S. Stanley III, D. Burkhardt, D. van Dijk, G. Wolf, and S. Krishnaswamy, “Manifold learning-based methods for analyzing single-cell rna-sequencing data,” *Current Opinion in Systems Biology*, vol. 7, pp. 36–46, 2018.
- [9] A. A. Galitsyna and M. S. Gelfand, “Single-cell hi-c data analysis: safety in numbers,” *Briefings in bioinformatics*, vol. 22, no. 6, p. bbab316, 2021.
- [10] Y. Zhang and G. Li, “Advances in technologies for 3d genomics research,” *Science China Life Sciences*, vol. 63, pp. 811–824, 2020.
- [11] S. V. Ulianov and S. V. Razin, “The two waves in single-cell 3d genomics,” in *Seminars in Cell & Developmental Biology*, vol. 121. Elsevier, 2022, pp. 143–152.
- [12] L. Wen and F. Tang, “Recent advances in single-cell sequencing technologies,” *Precision clinical medicine*, vol. 5, no. 1, p. pbac002, 2022.