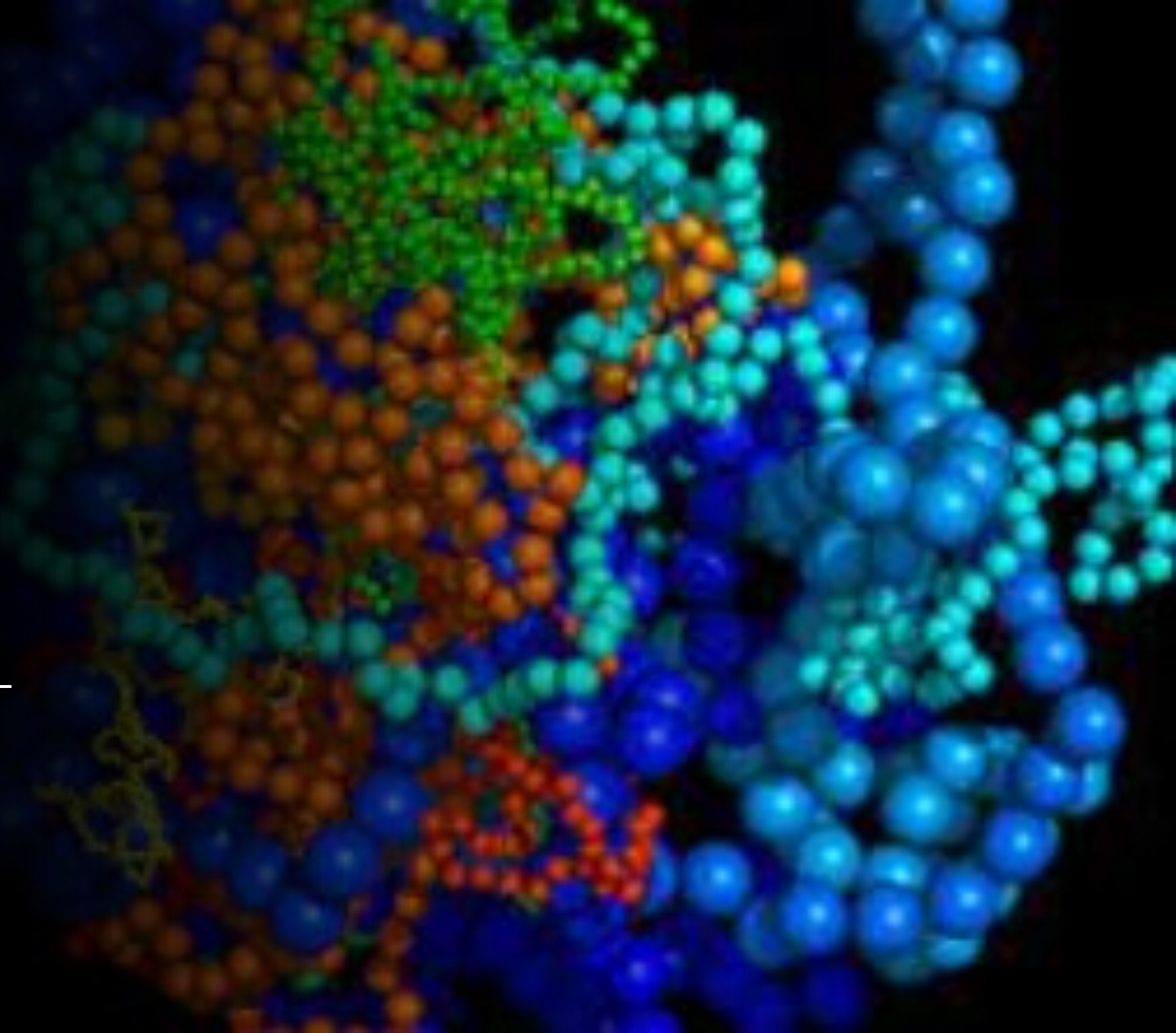


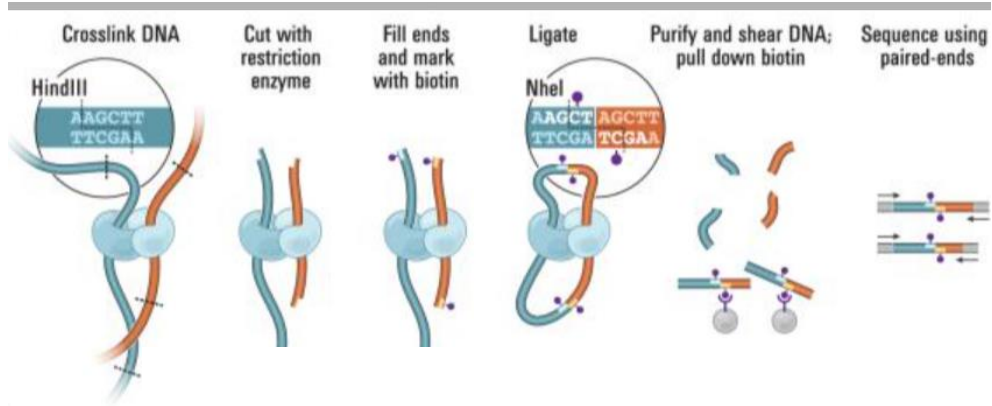


Extracting structural folding pattern of chromatin using chromatin condensation data

Group 5, Samira Mali

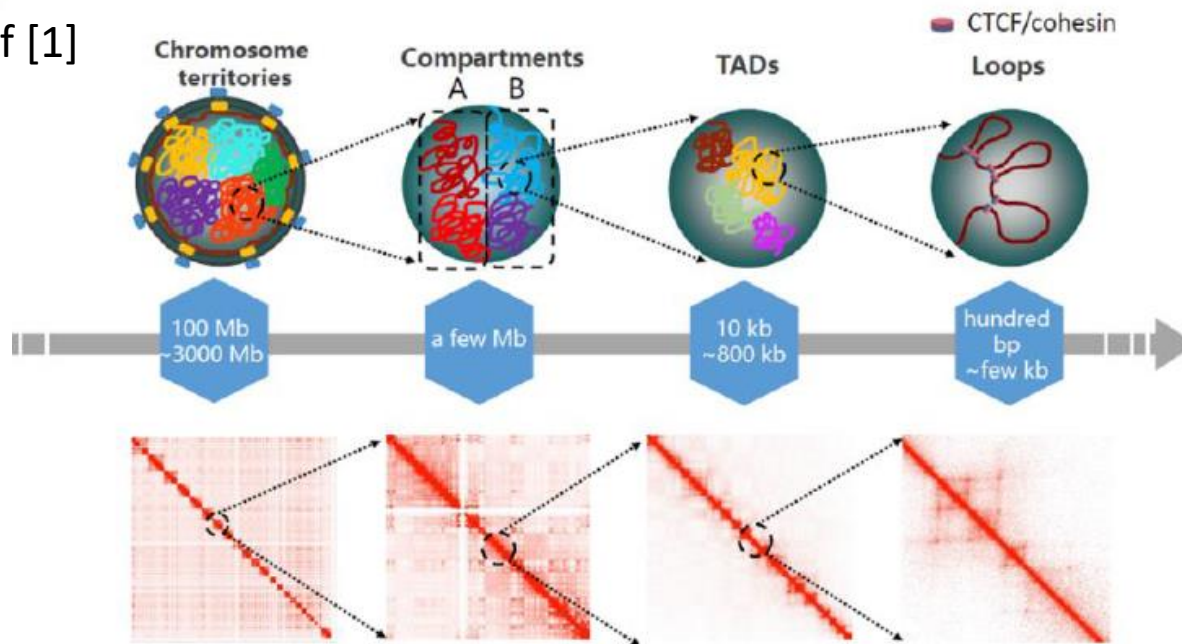


Introduction

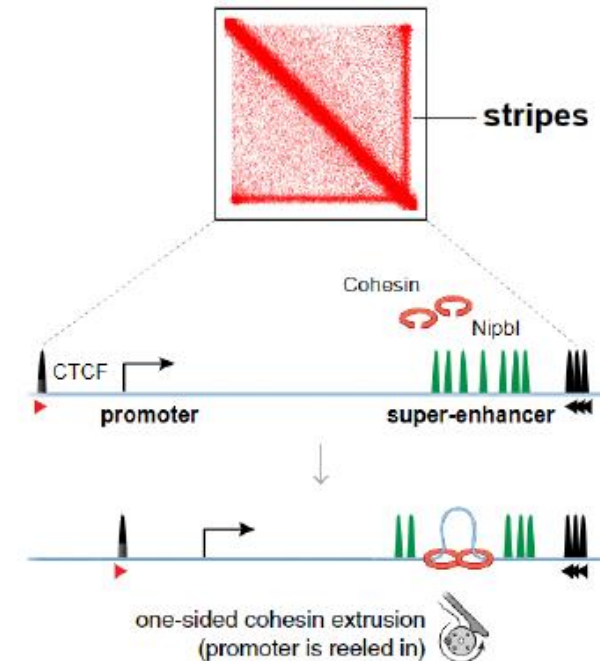


- Hi-C experiment
- In these Heatmaps, most of the chromatin folding patterns in the nucleus can be observed.
- Topologically Associating Domains (TAD), A/B compartments, loops, and stripes are some of the most well-known structural patterns.

From Ref [1]



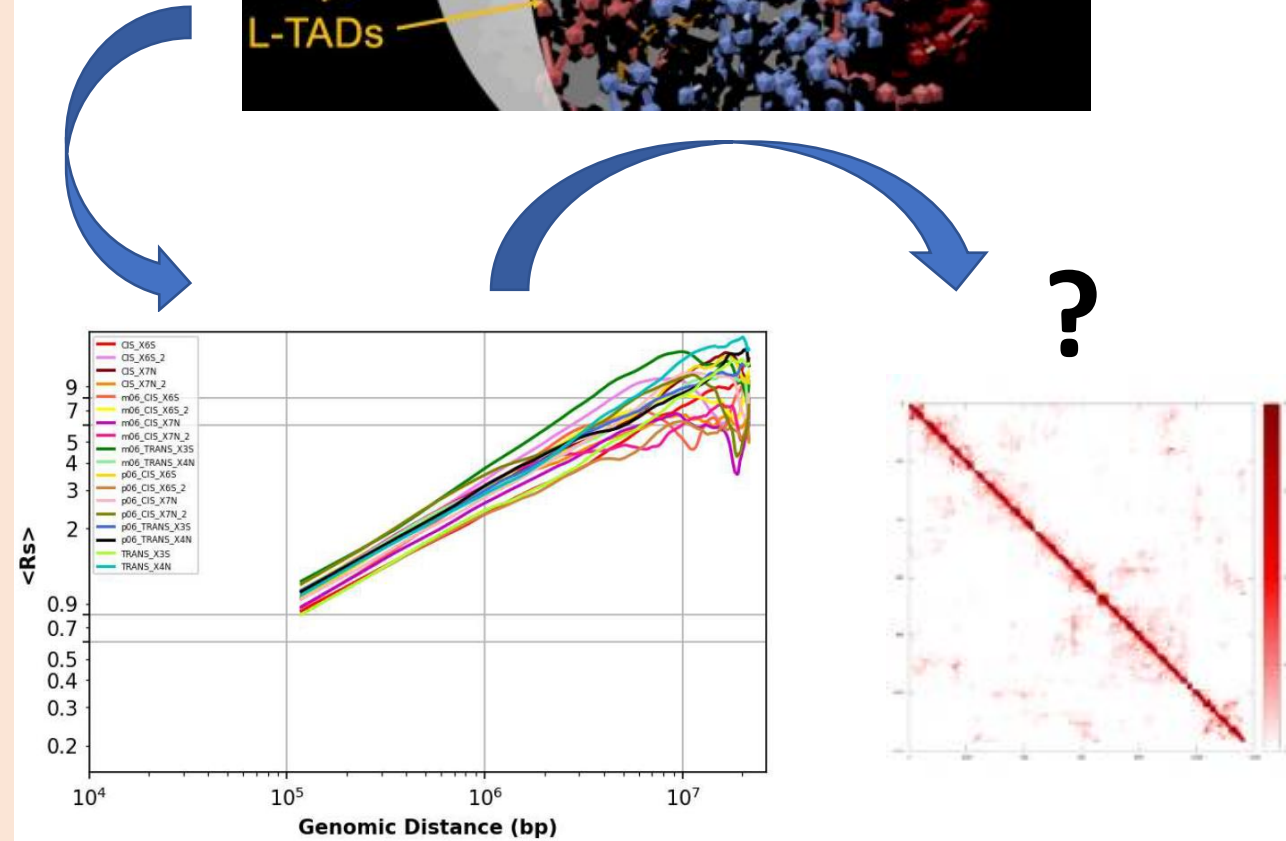
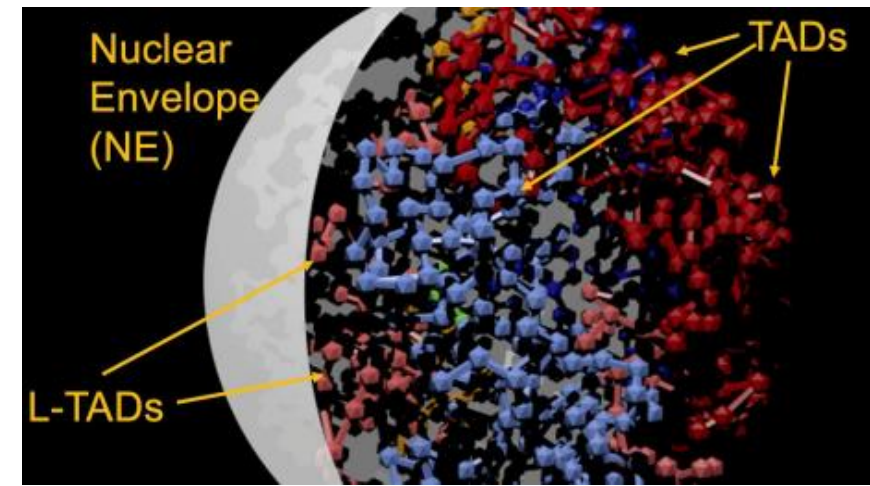
Courtesy from Qiu Sun



From Ref [14]

Problem

- In beads-on-a-string models of chromatin, the curves which show the dependence of the Euclidean spatial inter-particle distance of chromatin chains on the genomic distance play an important role to reveal the information of condensation of chromatin across whole genome.
- We have a computational modeling for chromatin in drosophila genome based on MD simulation[2] which we can extract these curves from the distance matrices of position of beads on ensemble of chromatin chain.
- *What I am interested in is whether we can obtain Hi-C folding patterns following training the data with CNN?*



Each curve represents one specific trajectory in the MD simulation.

- Provide data set as the symmetric square matrices
- We have 18 trajectories that show different chromosome configurations in Drosophila. In molecular dynamic simulation of chromatin chains in the Drosophila genome, we have 400K different configurations that correspond to 400K different timesteps for each trajectory. [2]

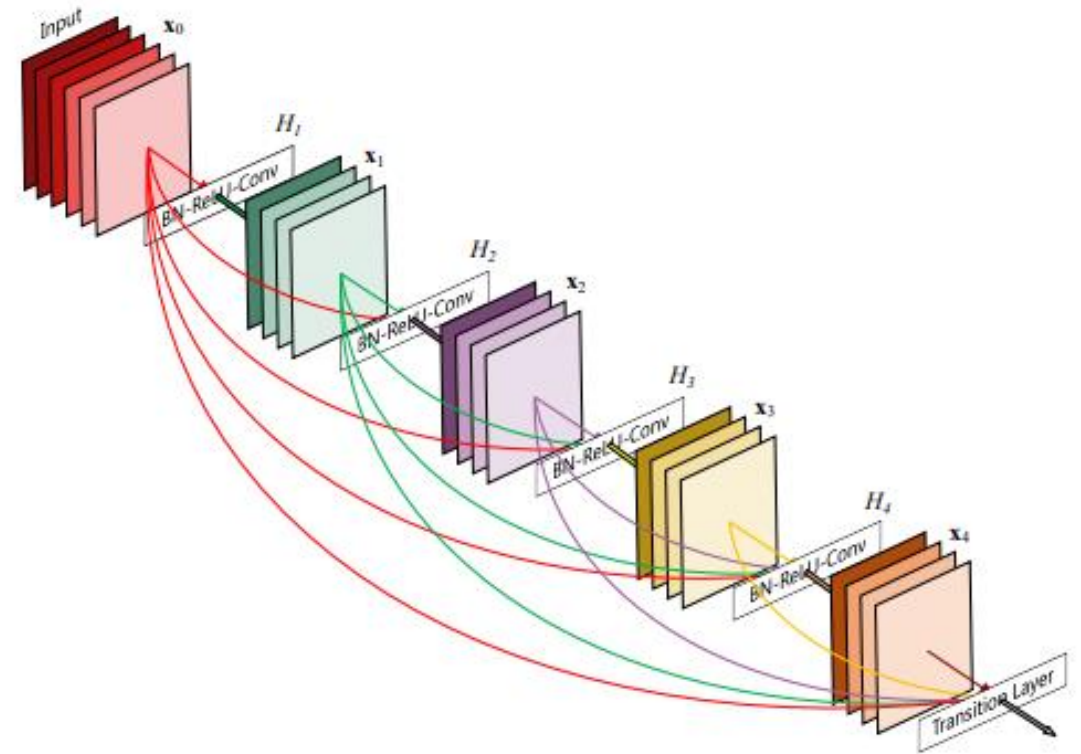


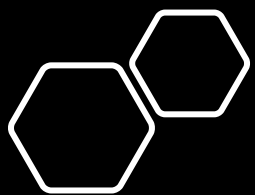
Milestone 1

dataset

Milestone 2

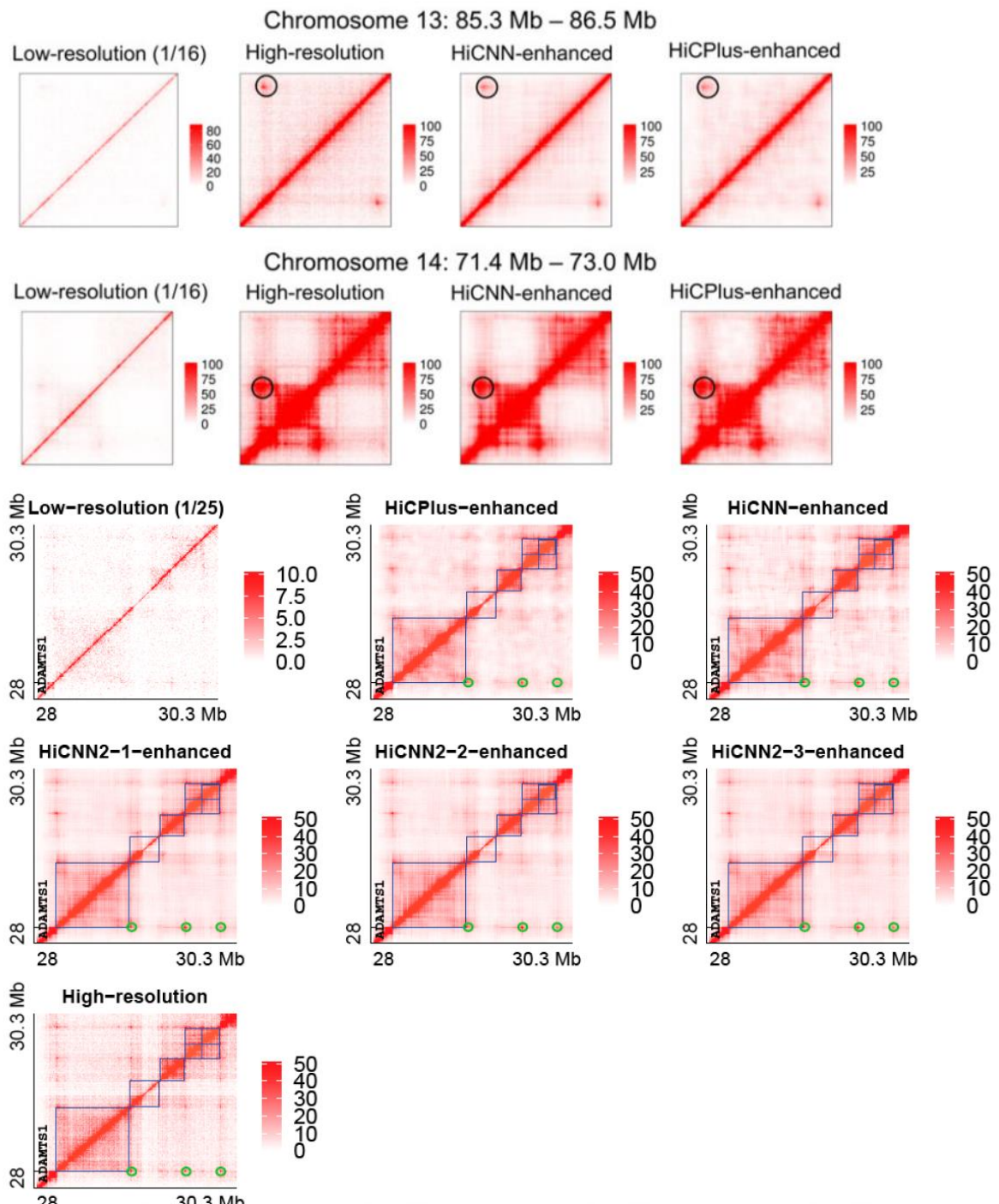
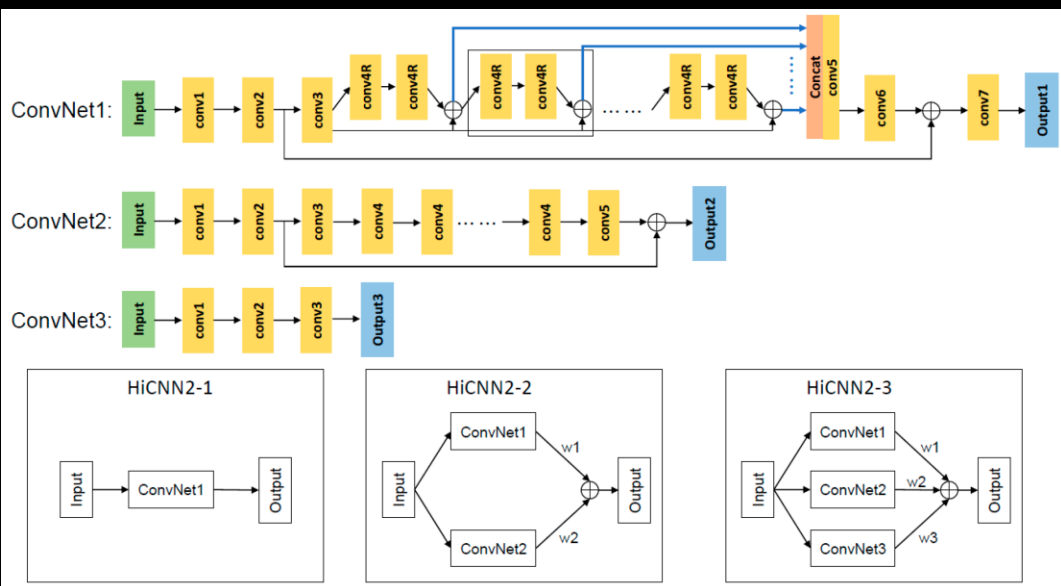
- Training the 2D matrices using **Convolutional Neural Networks** (ImageNet[3], LeNet [4] or AlexNet[5]).
- Training the 2D matrices using one of VGG[6], InceptionNet[7, 8, 9] ResNet[10] and DenseNet[11].





Milestone 3

Increasing the resolution
using HiCNN[12] or HiCNN2[13] approaches



Milestone 4

- **Metrics for evaluation:**

Comparing the result's mean squared error and Pearson's correlation coefficients with experimental high-resolution Hi-C data

References

- [1] Lieberman-Aiden, Erez, Nynke L. Van Berkum, Louise Williams, Maxim Imakaev, Tobias Ragozy, Agnes Telling, Ido Amit et al. "Comprehensive mapping of long-range interactions reveals folding principles of the human genome." *science* 326, no. 5950 (2009): 289-293.
- [2] Tolokh, Igor S., Nicholas Allen Kinney, Igor V. Sharakhov, and Alexey V. Onufriev. "Strong interactions between highly-dynamic lamina-associated domains and the nuclear envelope stabilize the 3D architecture of *Drosophila* interphase chromatin." *bioRxiv* (2022).
- [3] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition, pages 248–255. Ieee, 2009.
- [4] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
- [5] Yann LeCun, Bernhard E Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne E Hubbard, and Lawrence D Jackel. Handwritten digit recognition with a back-propagation network. In *Advances in neural information processing systems*, pages 396–404, 1990.
- [6] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [7] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander A Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *Thirty-first AAAI conference on artificial intelligence*, 2017.
- [8] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.
- [9] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016.
- [10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [11] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.
- [12] Liu, Tong, and Zheng Wang. "HiCNN: a very deep convolutional neural network to better enhance the resolution of Hi-C data." *Bioinformatics* 35, no. 21 (2019): 4222-4228.
- [13] Liu, Tong, and Zheng Wang. "HiCNN2: enhancing the resolution of Hi-C data using an ensemble of convolutional neural networks." *Genes* 10, no. 11 (2019): 862.
- [14] Vian, Laura, Aleksandra Pękowska, Suhas SP Rao, Kyong-Rim Kieffer-Kwon, Seolkyoung Jung, Laura Baranello, Su-Chen Huang et al. "The energetics and physiological impact of cohesin extrusion." *Cell* 173, no. 5 (2018): 1165-1178.
- [15] Sun, Qiu, Alan Perez-Rathke, Daniel M. Czajkowsky, Zhifeng Shao, and Jie Liang. "High-resolution single-cell 3D-models of chromatin ensembles during *Drosophila* embryogenesis." *Nature communications* 12, no. 1 (2021): 1-12.