"Exploration of Hi-C patterns through computer simulations"

By: Elizaveta Rybnikova

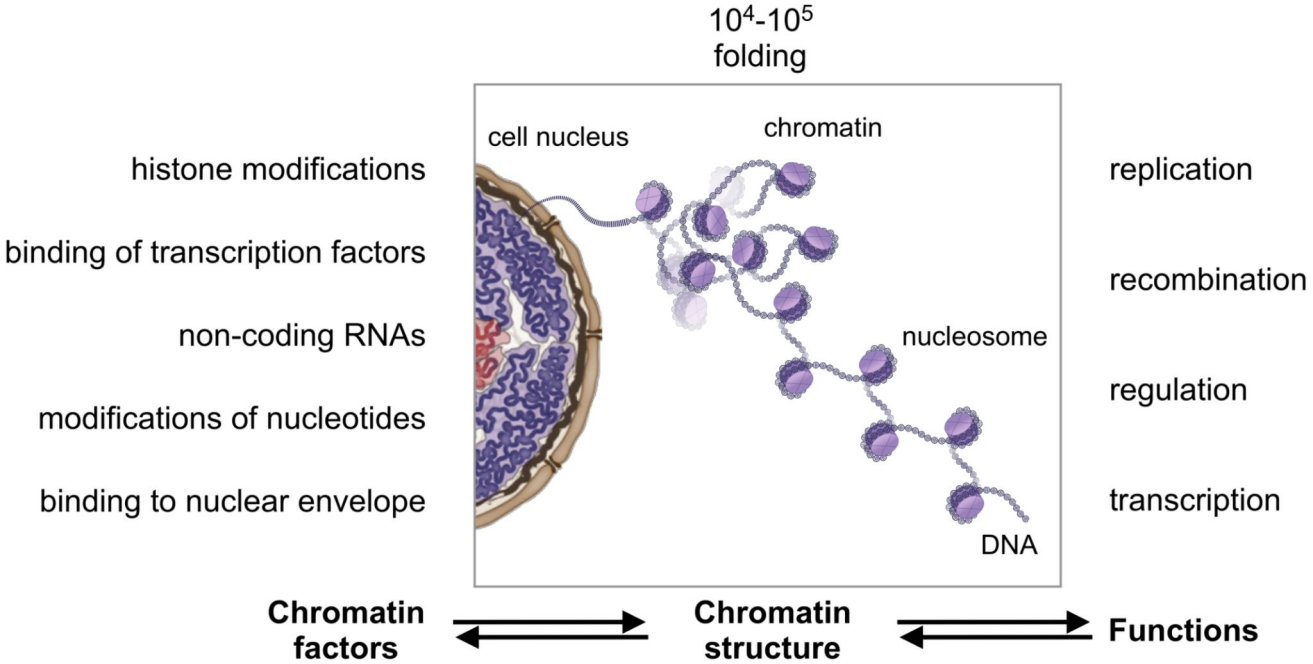
Program: MIT PRIMES 2023

Mentors: Henrik Pinholt, Dr. Aleksandra Galitsyna

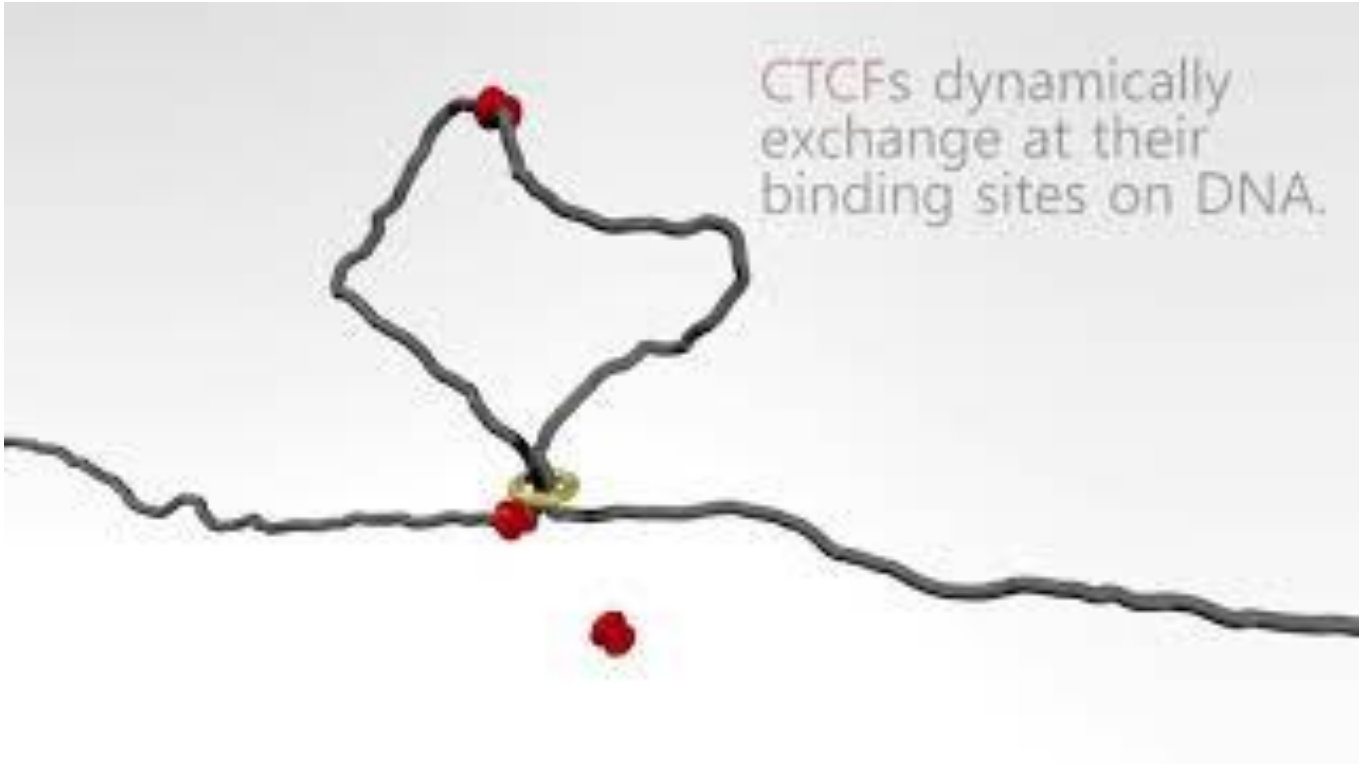
Mirny Lab

DNA organisation is very complex

Eukaryotic nucleus:



Loop extrusion mechanism



Experimental method to analyse: Hi-C

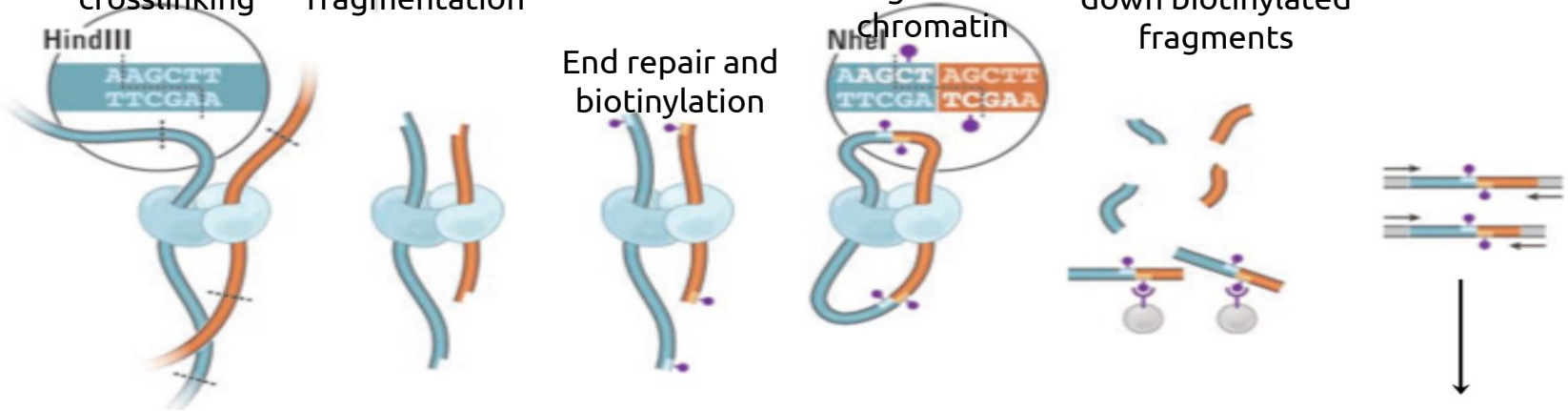
“FIX”:
Chromatin crosslinking

“CUT”:
chromatin fragmentation

End repair and biotinylation

“LIGATE”:
Ligate fragmented chromatin

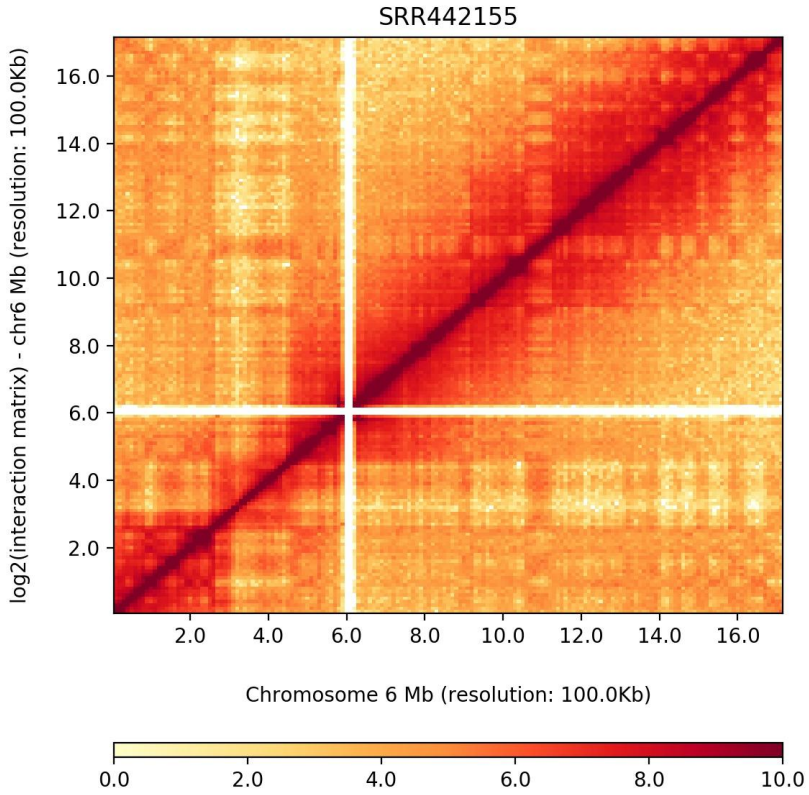
“ENRICH”:
shear and pull down biotinylated fragments



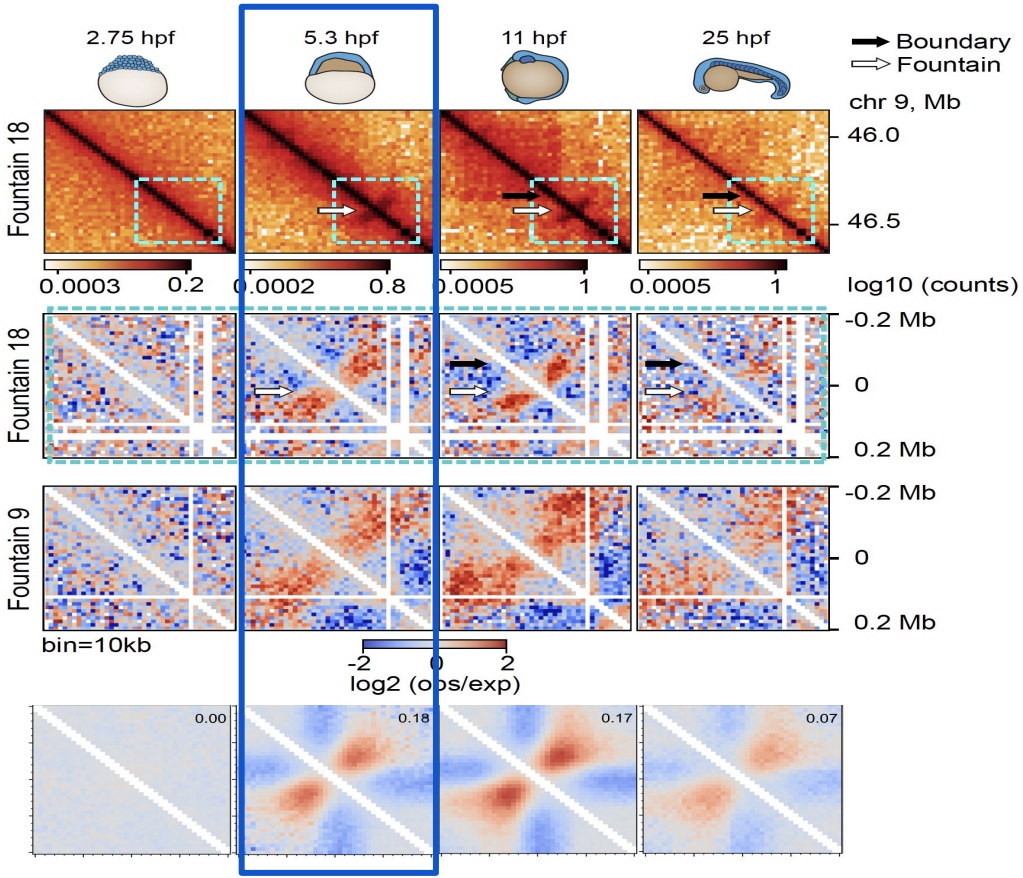
“SEQUENCE”:
Paired-end sequencing of biotinylated fragments

Lieberman-Aiden et al. 2009

Hi-C map example



New structures: **fountains**

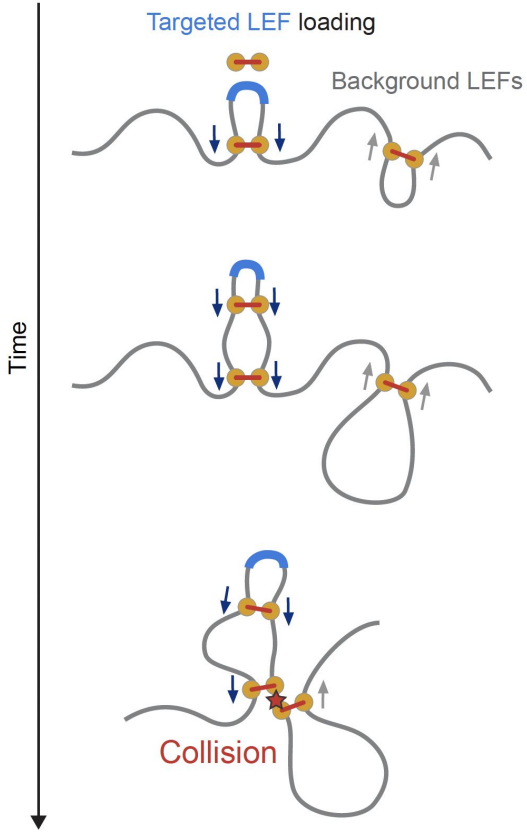


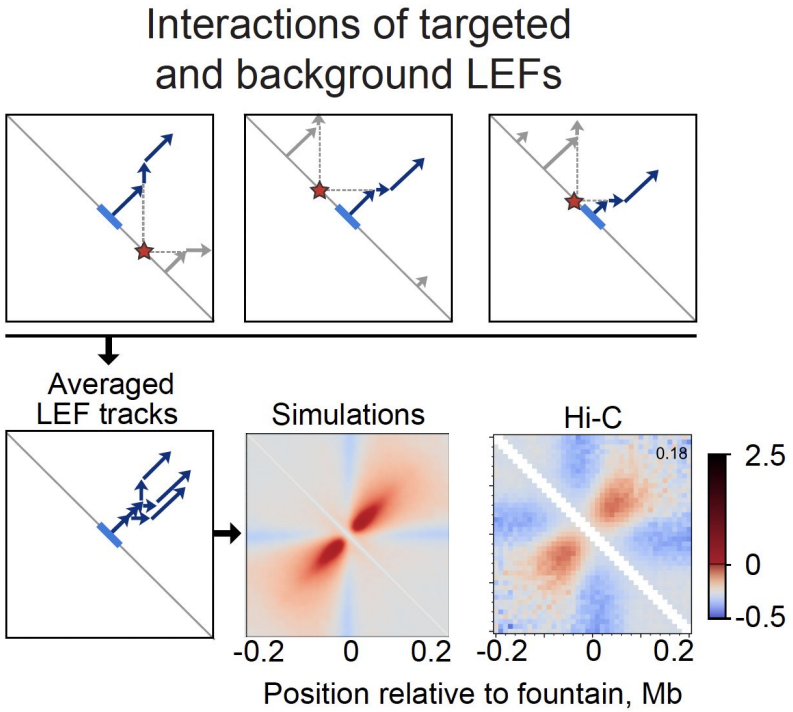
Example 1

Example 2

Average fountain

Mechanism of **fountain formation:** loop extrusion





Project goals

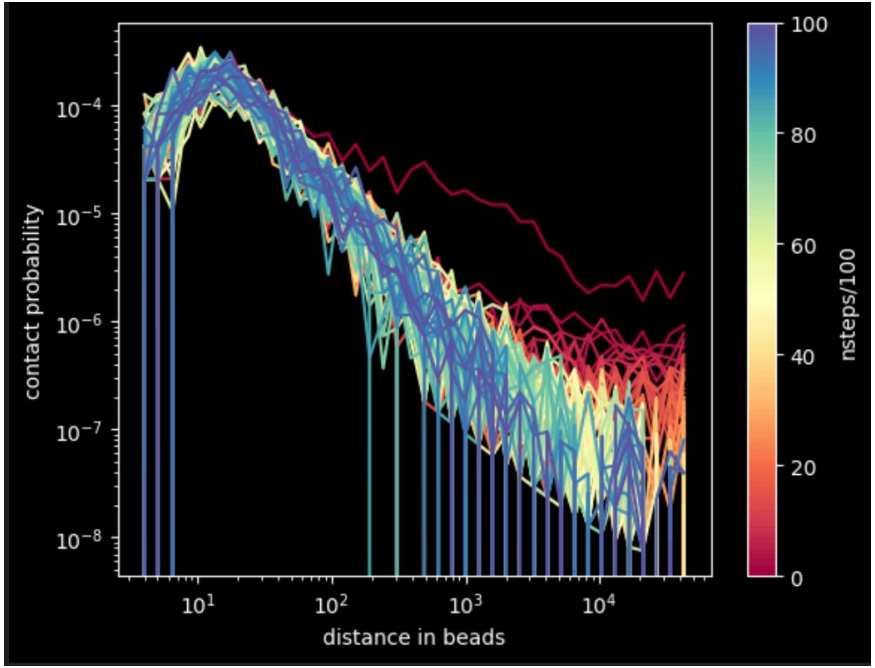
- ***Goal:*** *Understanding mechanisms of formation of different chromatin structures visible in Hi-C through simulations of loop extrusion with changing parameters.*
- ***Aim 1:*** *Simulate fountains and characterise how their shape depends on simulations and, in particular, extrusion parameters.*
- ***Aim 2****: Compare simulated and Hi-C fountains and create a method to infer loop extrusion parameters from fountain shapes.*
- ***Aim 3 and further****: With the developed methods, simulate TADs, stripes and loops, and infer the loop extrusion parameters for them.*

Aim 1: Simulation's work

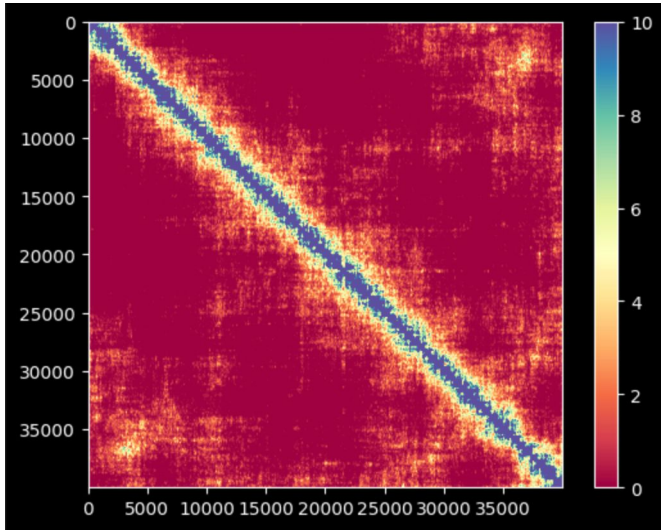
- First, we model all of the cohesin's movement and activity in one dimension, representing the DNA as a a list with indicated positions of CTCFs and cohesins. On this step we change the parameters for cohesin's activity which affects the size and placement of the loop and, consequently, fountains.
- After that we use the one dimensional simulation to create a 3D model of the polymer by placing cohesins on their indicated places and adding forces affecting the polymer structure.
- We ran all of our simulations for 10,000 steps, which we estimated to be a value that will bring us enough data for the Hi-C map creation.

Aim 1, results: Simulations of polymer as a base for future simulation of loop extrusion.

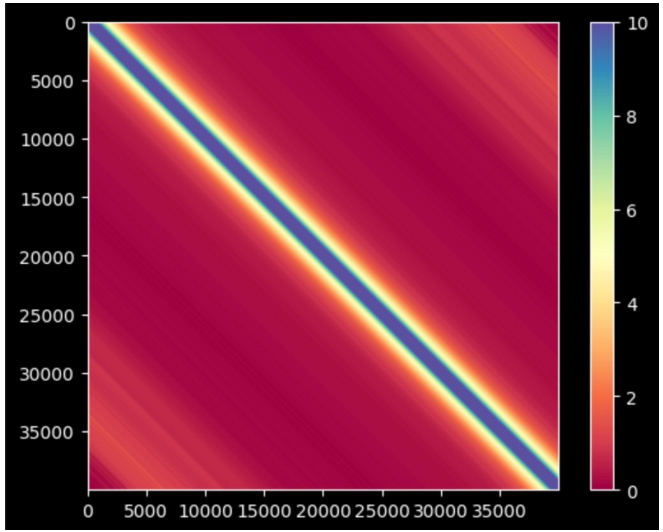
Dependence of contact probability on the distance between beads in simulation. Colour represents sequential steps of simulations, and the time course represents the equilibration of the polymer.



Aim 1, results: Simulations of loop extrusion (Hi-C maps)

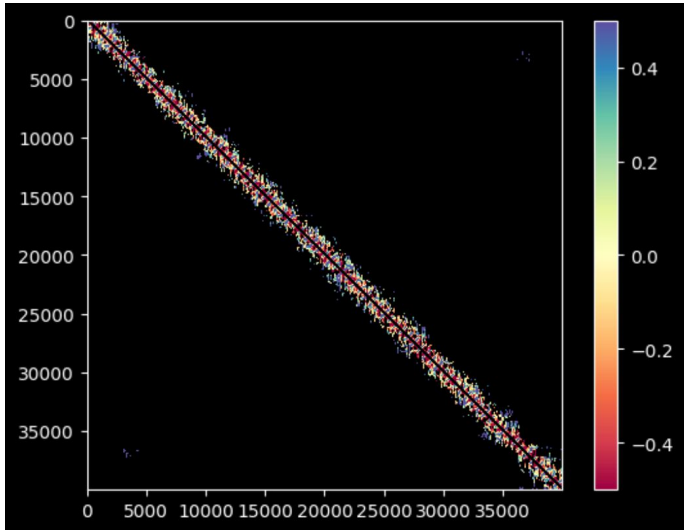


The original Hi-C map from a matrix directly from the simulation.

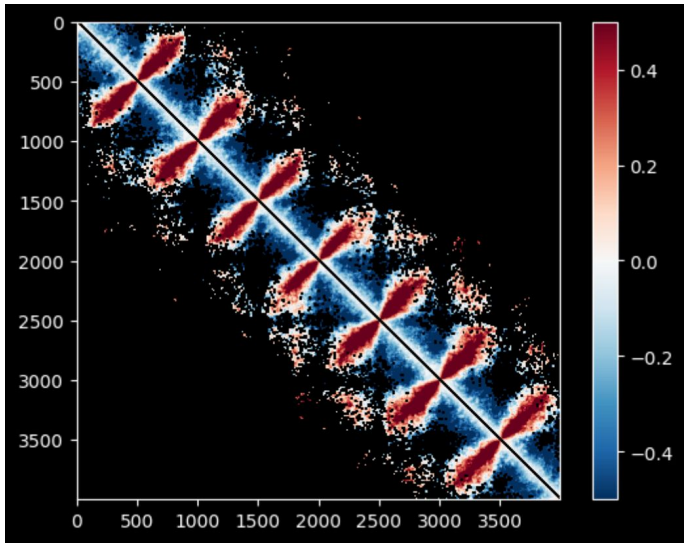


Hi-C map after averaging every diagonal on it.

Aim 1, results: Simulations of loop extrusion (Hi-C maps)



Hi-C map after performing a logarithmic distribution algorithm

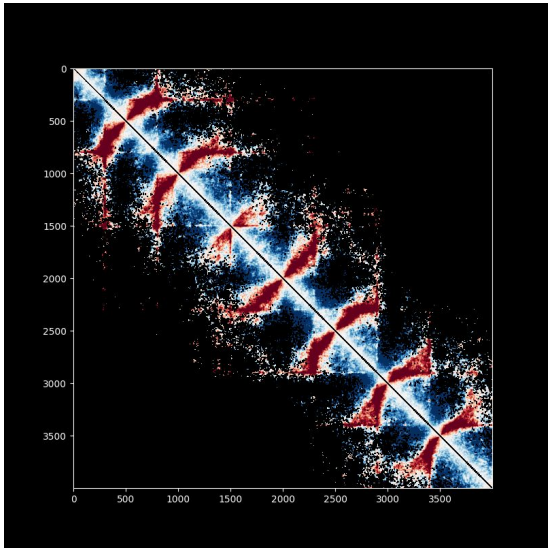


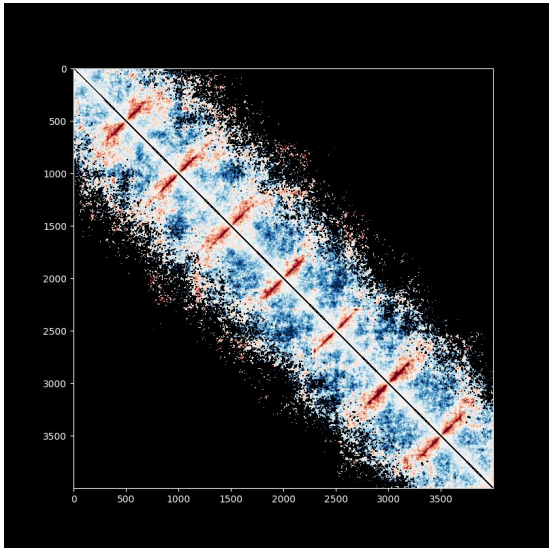
Hi-C map with fountains concentrated at target sites.

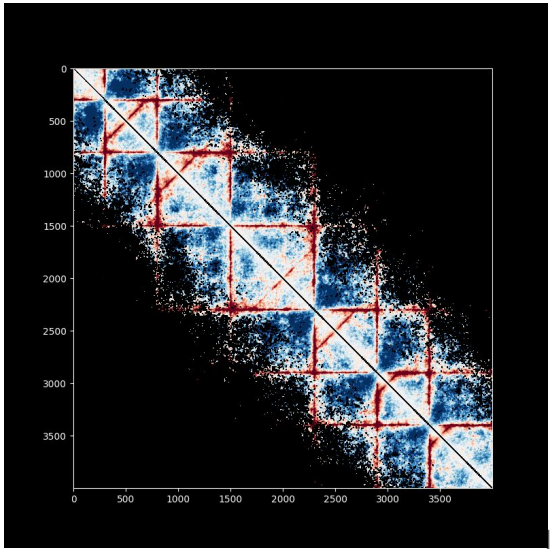
Results

We created several hundreds of Hi-C maps with different parameters of cohesin and target sites' attraction for further creation of the dataset for the neural network.

Here are some of the most distinct types of maps.

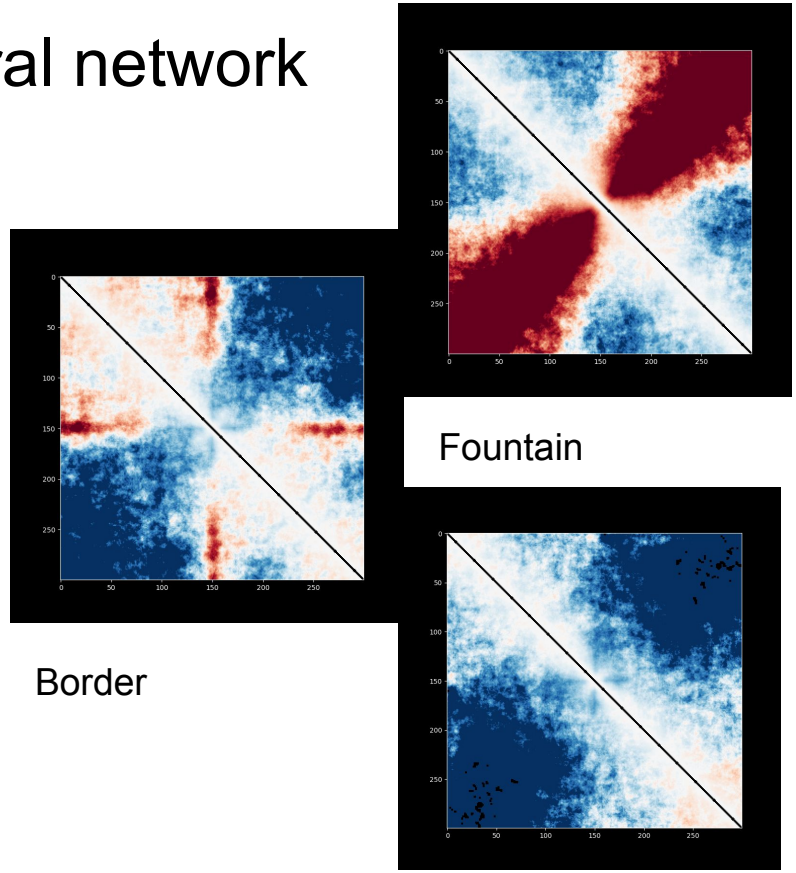






Aim 2, start: Dataset for the neural network

- For the first steps in training our neural network we create a dataset with three classes of Hi-C features: fountains, borders and control group with no features.
- Our initial network will learn to distinguish between these three classes, so then we can try to make it distinguish between different types of fountains.
- Now the network's accuracy is below 0.5, which means that we need to expand our dataset.



Fountain

Border

No features

Aim 2: Next steps for the neural network

- Expand the dataset to include more distinguisheble types of fountains.
- Add layers of complexity to the neural network, also changing the input layer to analyse

Next plans

- Expand the Hi-C dataset for the network in order to train it to predict the parameters for cohesins used.
- Get real life parameters from the network.
- Simulate a map on these parameters, then use convolutional neural network to find fountains on real Hi-C maps.
- Fountains are known to be on the enhancers in the DNA (i.e. regulatory fragments of DNA that control gene expression), so we can prove it (or not prove it) with our collected data.

Acknowledgements

- I would like to thank my mentors Dr. Alexandra Galitsyna and Henrik Pinholt for introducing me to this field and guiding my project.
- I would also like to thank MIT PRIMES and MIT Mirny Lab for making this project possible.