

# Stitching Chromatin Puzzle with Hi-C

Yiftach Kolb

April 17, 2021

## 1 Main Assumptions and Ideas

- Two blocks of the Hi-C matrix that should be aligned together create a pattern that 'looks right', and when they shouldn't be adjacent or adjacent in the wrong orientation/side it 'looks wrong'.
- If we could indeed say whether every pair should(n't) be adjacent, we can patch together the chromosome using a greedy algorithm. This turns reduces the problem from a hard TSP situation to an  $O(n^2)$ .
- How do we achieve step 1? We can come up with a good scorign scheme to serve as a distance 'metric' or something. But probably a better way is to train a classifier.
- The Classifier. Maybe any classifier can do the job but remember that Hi-C matrices are basically images  $\Rightarrow$  NN or some similar ML technique are very good at image classification.
- What do we train/test on? Take HI-C that that is correct, cut it to pieces. We know the order of the pieces so we can train a classifier to predict if two pieces belong.
- Basically we rearrange blocks of the Hi-C matrix. Cut portions (probably along the diagonal) and feed this to the NN to train on.

1.a

## 2 Reference

List of Figures

List of Tables