Hi-C interaction matrix correction using ICE in Rust

Bachelor thesis defense

Felix Karg

13. Juli 2019

University of Freiburg

Content

Content

Enhancer Promoter Interaction

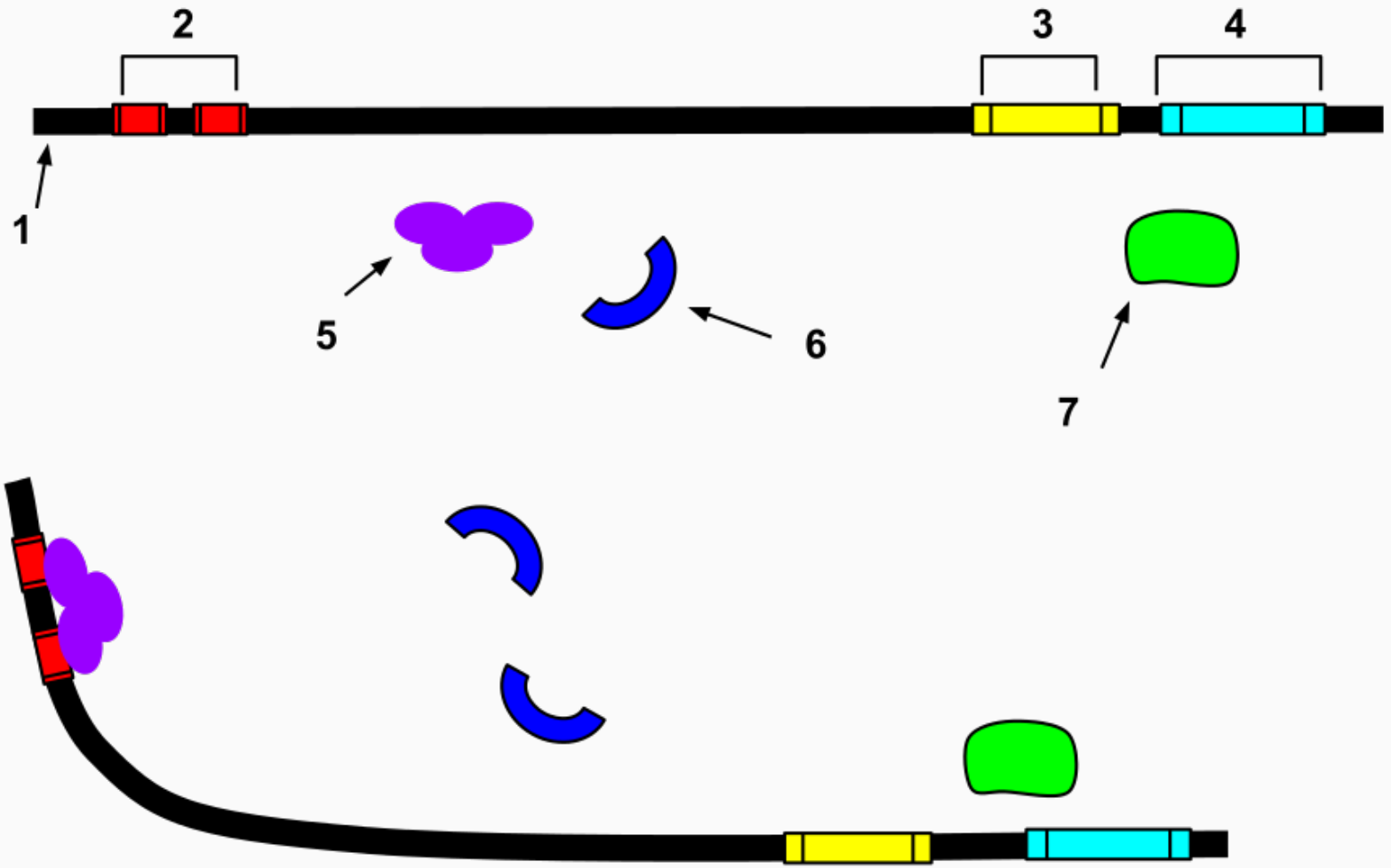

Image adapted from [1].

Enhancer Promoter Interaction

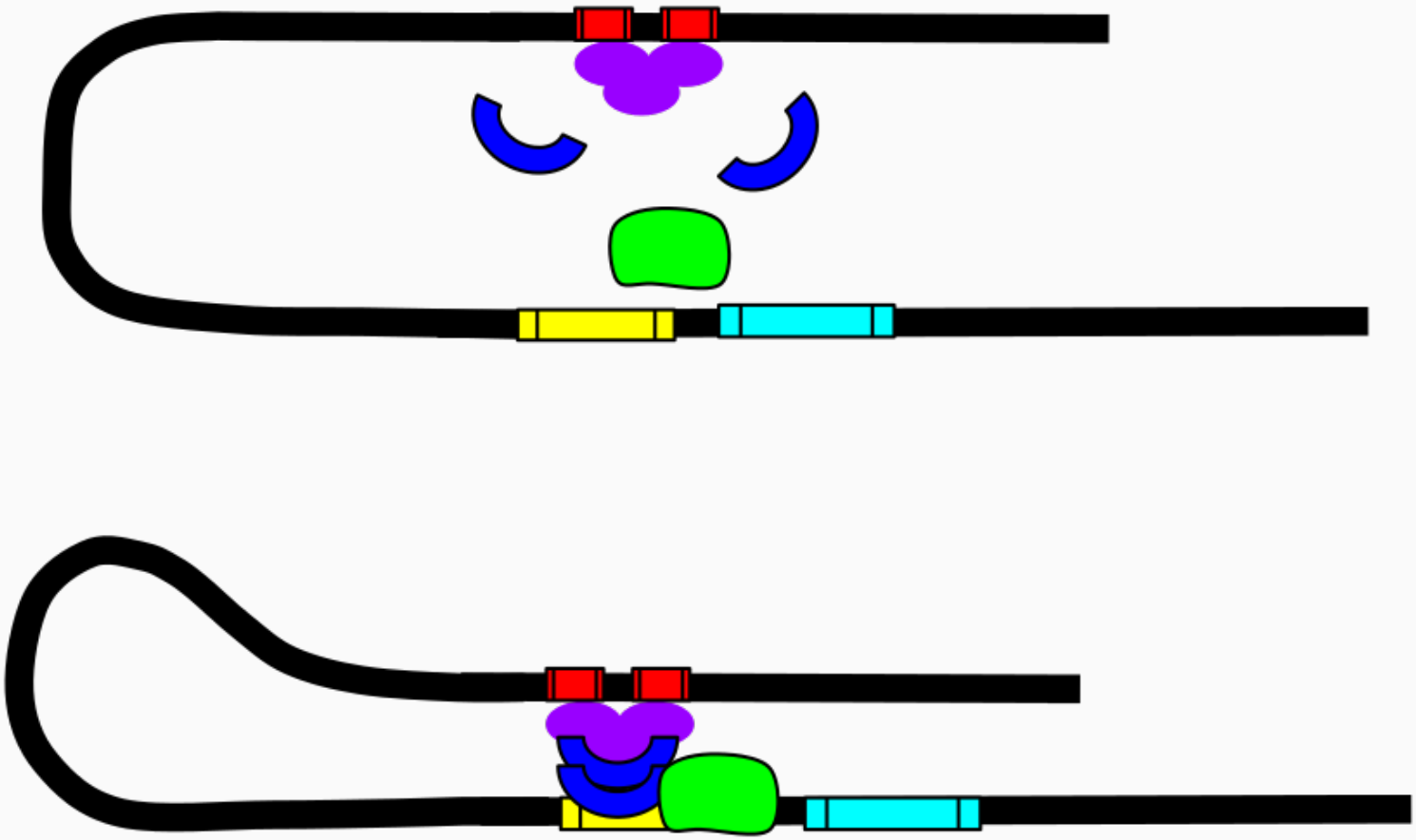

Image adapted from [1].

Chromatin

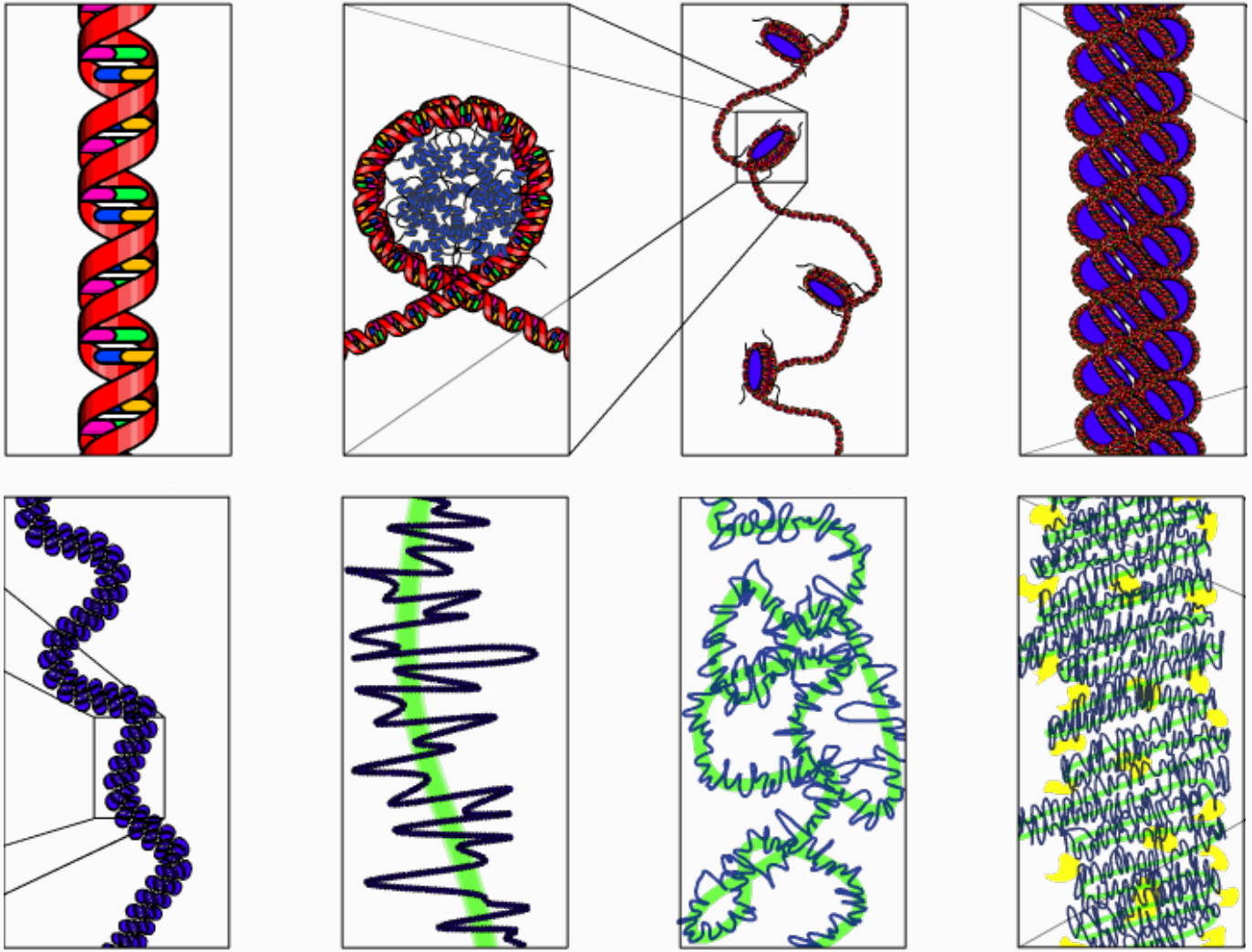

Image adapted from [2].

Spatial Structure

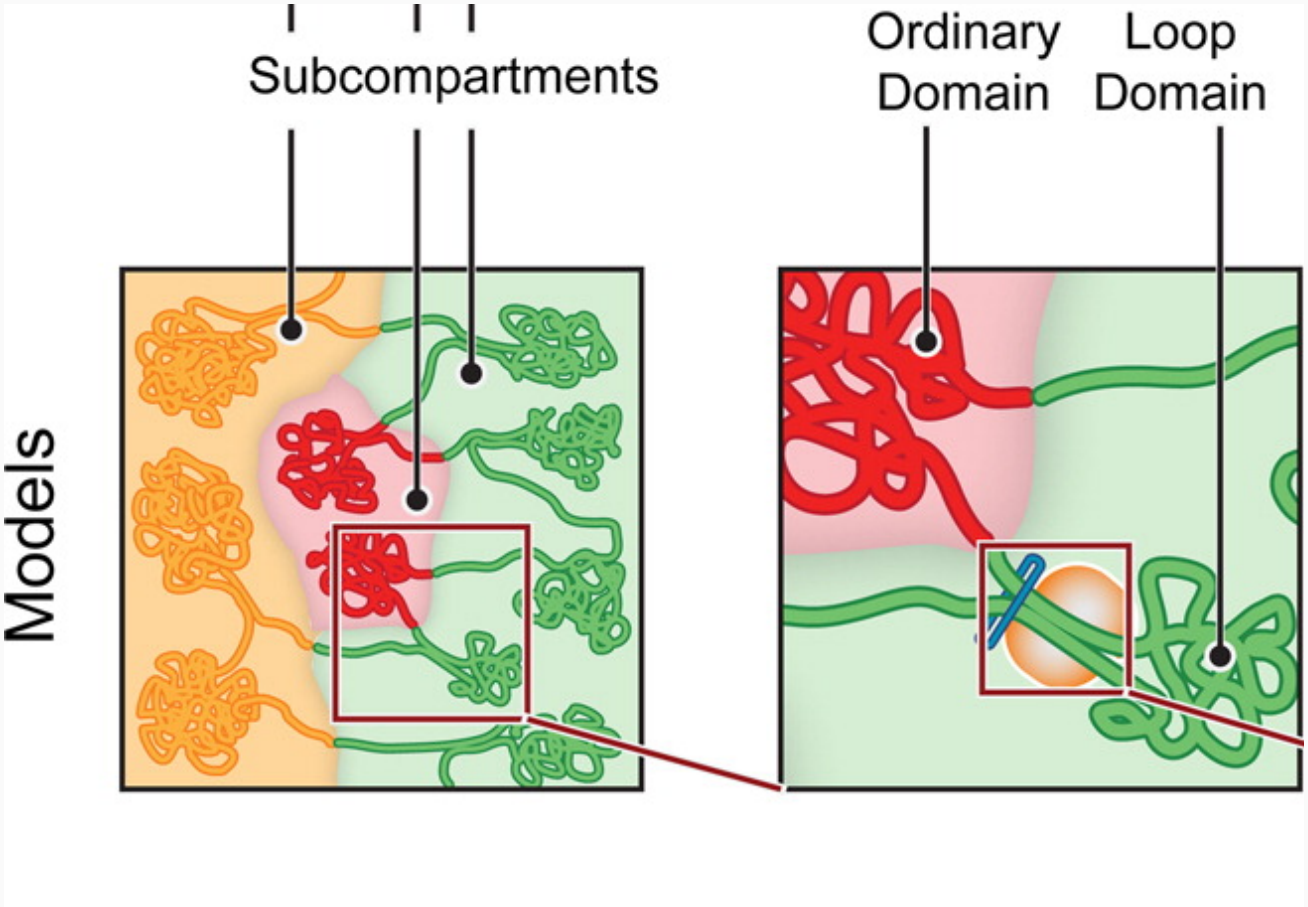

Image adapted from [3].

Enhancers may act on genes from other chromosomes [4]!

Content

High-Throughput 3C (Hi-C)

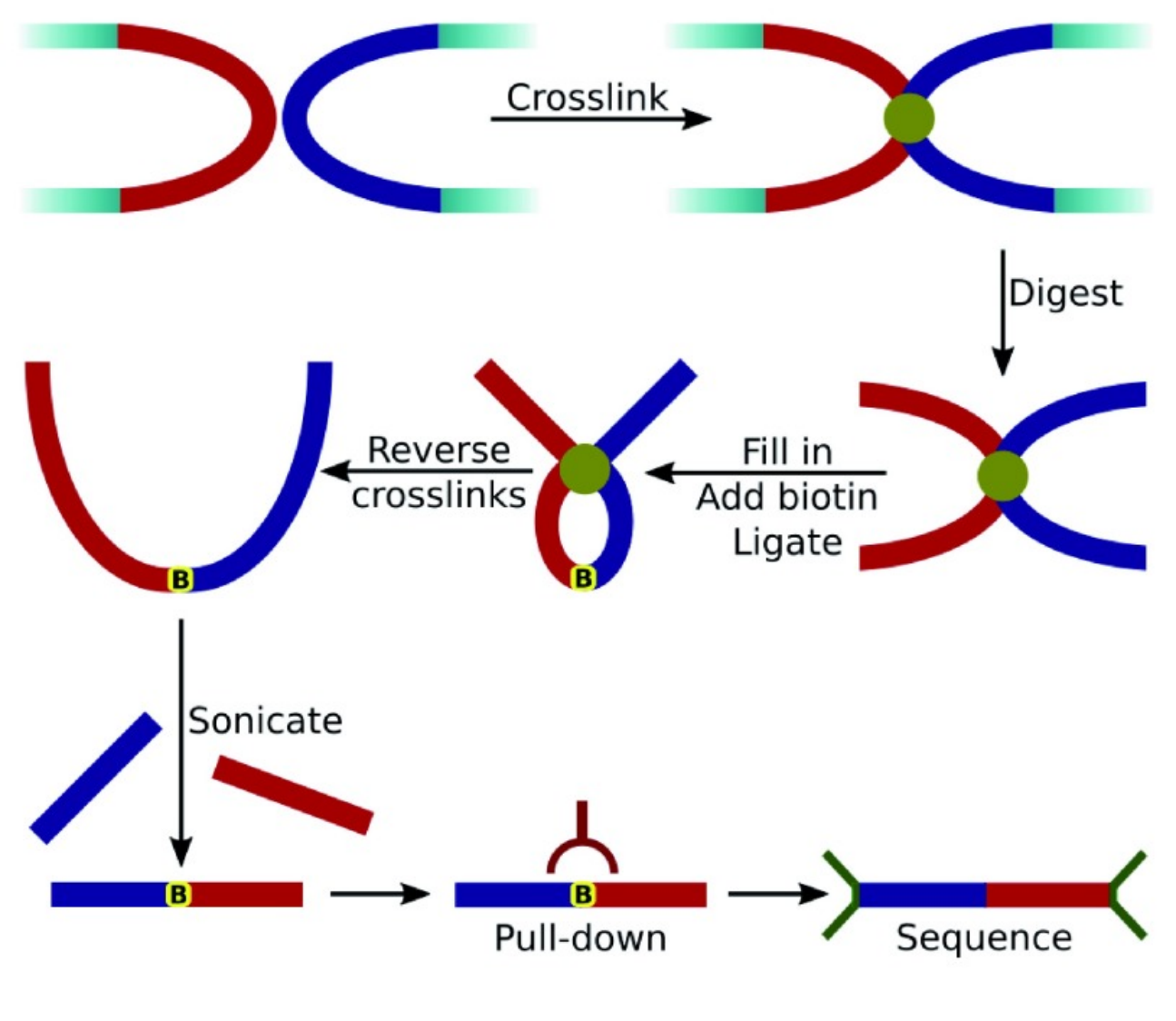

Image adapted from [5].

High-Throughput 3C (Hi-C)

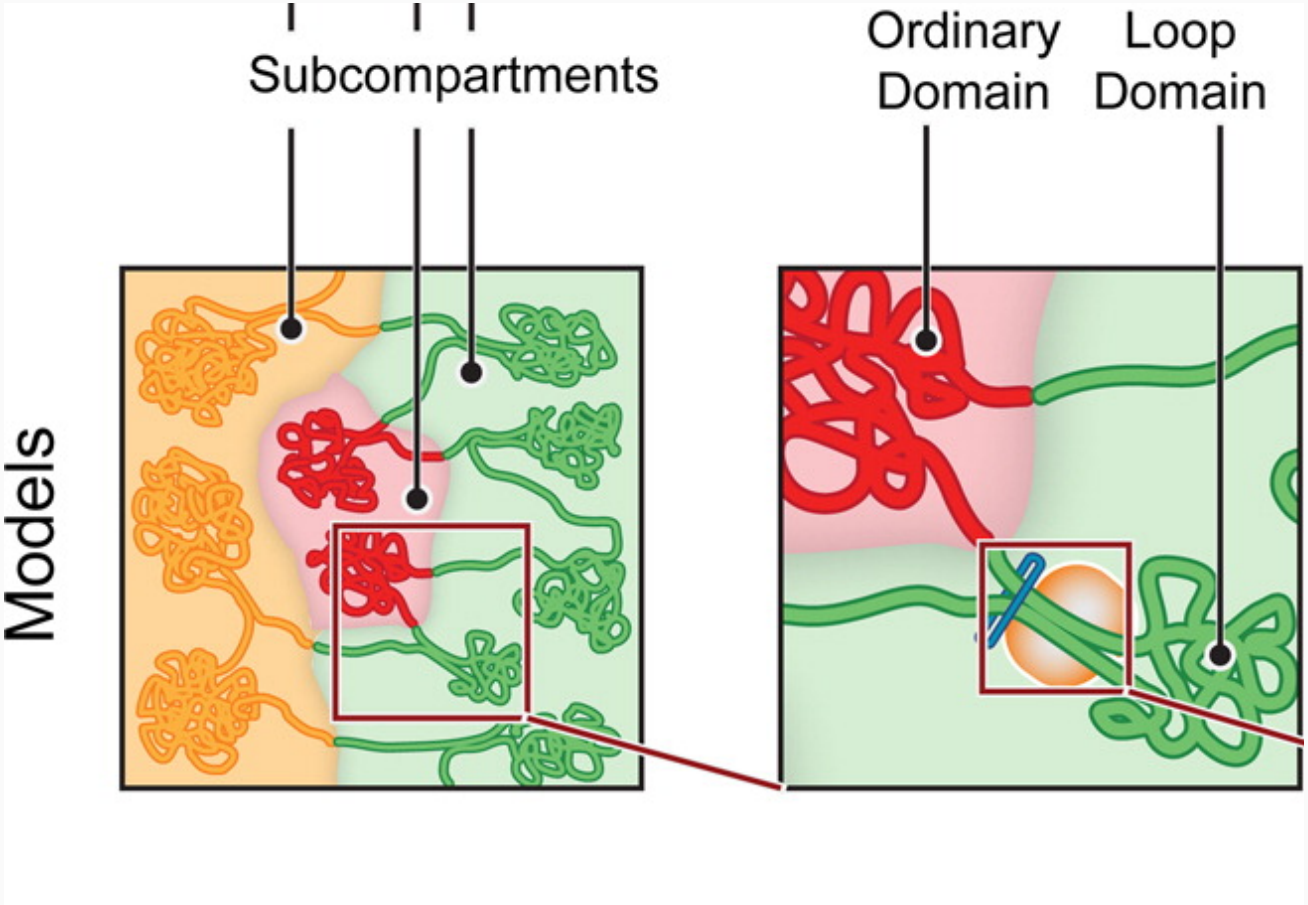

Image adapted from [3].

High-Throughput 3C (Hi-C)

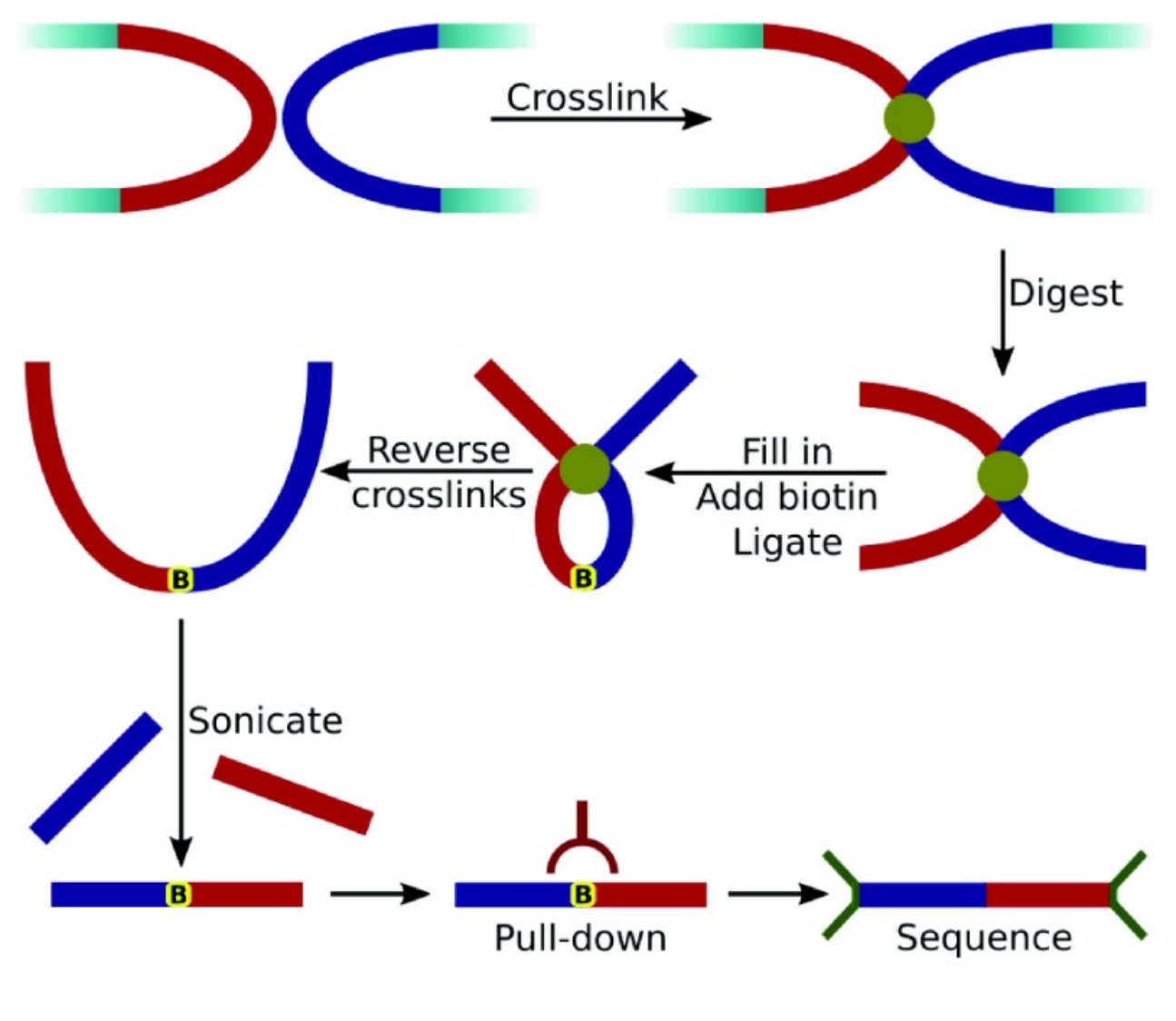

Image adapted from [5].

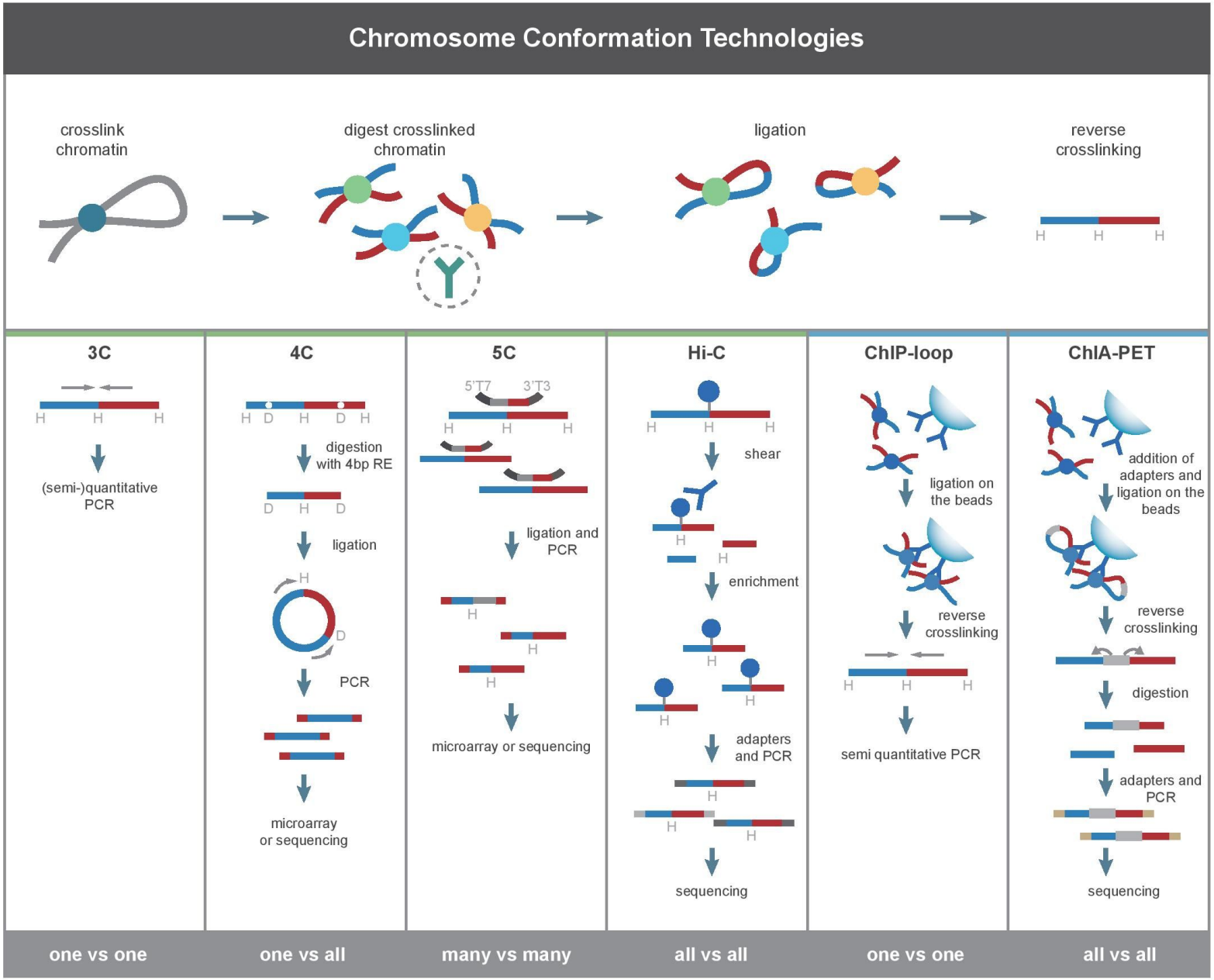
Chromosome Conformation Technologies
crosslink chromatin
digest crosslinked chromatin
ligation
reverse crosslinking
3C
(semi-)quantitative PCR
4C
digestion with 4bp RE
ligation
PCR
microarray or sequencing
5C
ligation and PCR
microarray or sequencing
Hi-C
shear
enrichment
adapters and PCR
sequencing
ChIP-loop
ligation on the beads
reverse crosslinking
semi quantitative PCR
ChIA-PET
addition of adapters and ligation on the beads
reverse crosslinking
digestion
adapters and PCR
sequencing
one vs one
one vs all
many vs many
all vs all
one vs one
all vs all

Image from [6].

HiCExplorer

HiCExplorer

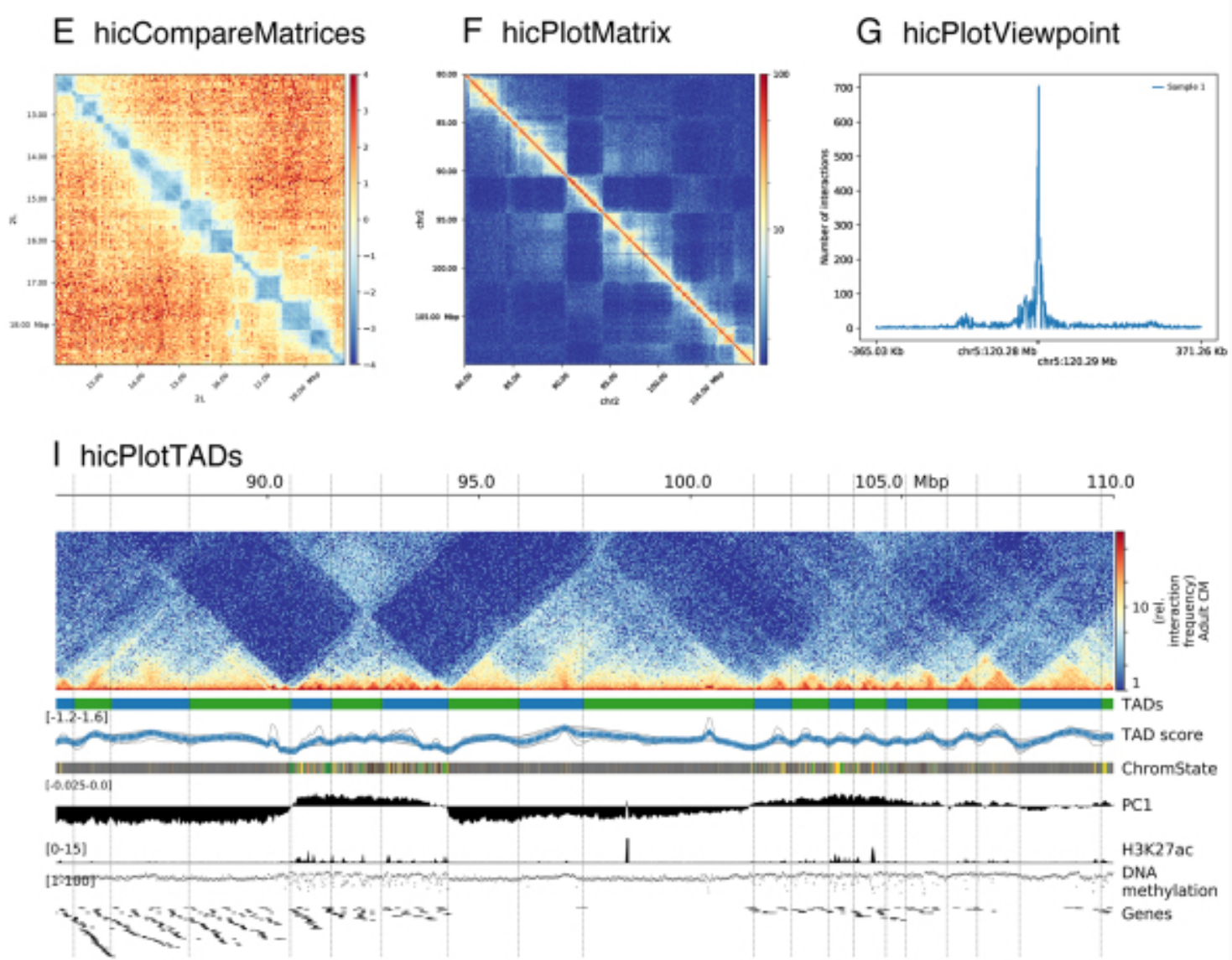

Image adapted from [7].

Content

Iterative Correction and Eigenvector decomposition

Algorithm as described in [8].

Iterative Correction and Eigenvector decomposition

Iterative Correction and Eigenvector decomposition

- O_{ij}: raw data

Iterative Correction and Eigenvector decomposition

- O_{ij}: raw data

- T_{ij}: relative contact probabilities

Iterative Correction and Eigenvector decomposition

- O_{ij}: raw data

- T_{ij}: relative contact probabilities
- W_{ij}: working copy of O_{ij}, becoming T_{ij}

Iterative Correction and Eigenvector decomposition

- O_{ij}: raw data

- B_i, B_j: cumulative biases
- T_{ij}: relative contact probabilities
- W_{ij}: working copy of O_{ij}, becoming T_{ij}

Iterative Correction and Eigenvector decomposition

- O_{ij}: raw data
- S_i: sum of row i of W_{ij}
- B_i, B_j: cumulative biases
- T_{ij}: relative contact probabilities
- W_{ij}: working copy of O_{ij}, becoming T_{ij}

Iterative Correction and Eigenvector decomposition

Goal: Obtain B and T_{ij}.

Iterative Correction and Eigenvector decomposition

Goal: Obtain B and T_{ij}.

Do this by explicitly solving:

$$O_{ij} = B_i B_j T_{ij} \tag{1}$$

$$\sum_{i=1,|i-j|>1}^{N} T_{ij} = 1 \tag{2}$$

Iterative Correction and Eigenvector decomposition

$$\forall_j \sum_{i=1}^{N} T_{ij} = 1 \tag{3}$$

$$\forall_i \sum_{j=1}^{N} T_{ij} = 1 \tag{4}$$

Iterative Correction and Eigenvector decomposition

Each iteration, solve:

Iterative Correction and Eigenvector decomposition

Each iteration, solve:

$$S_i = \sum_j W_{ij} \tag{5}$$

$$\Delta B_i = S_i / mean(S) \tag{6}$$

Iterative Correction and Eigenvector decomposition

Each iteration, solve:

$$S_i = \sum_j W_{ij} \tag{5}$$

$$\Delta B_i = S_i / mean(S) \tag{6}$$

$$W_{ij} = W_{ij} / \Delta B_i \Delta B_j \tag{7}$$

$$B_i = B_i \cdot \Delta B_i \tag{8}$$

Content

Content

Content

Sources i

- J. Cheff, "Enhancer nucleotide sequence." `https://commons.wikimedia.org/wiki/File:Enhancer_Nucleotide_Sequence.svg`, 2015. **Licensed under CC BY-SA 4.0; Image has been cropped; accessed 2019-07-12.**
- R. Wheeler, "Chromatin structures." `https://commons.wikimedia.org/wiki/File:Chromatin_Structures.png`, 2005. **Licensed under CC BY-SA 3.0; Image has been cropped; accessed 2019-06-26.**

Sources ii

- 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.
- C. G. Spilianakis, M. D. Lalioti, T. Town, G. R. Lee, and R. A. Flavell, "Interchromosomal associations between alternatively expressed loci," *Nature*, vol. 435, no. 7042, p. 637, 2005.

Sources iii

- S. Wingett, P. Ewels, M. Furlan-Magaril, T. Nagano, S. Schoenfelder, P. Fraser, and S. Andrews, "Hicup: pipeline for mapping and processing hi-c data," *F1000Research*, vol. 4, 2015.
- G. Li, L. Cai, H. Chang, P. Hong, Q. Zhou, E. V. Kulakova, N. A. Kolchanov, and Y. Ruan, "Chromatin interaction analysis with paired-end tag (chia-pet) sequencing technology and application," *BMC Genomics*, vol. 15, p. S11, Dec 2014.

Sources iv

- J. Wolff, V. Bhardwaj, S. Nothjunge, G. Richard, G. Renschler, R. Gilsbach, T. Manke, R. Backofen, F. Ramírez, and B. A. Grüning, "Galaxy hicexplorer: a web server for reproducible hi-c data analysis, quality control and visualization," *Nucleic acids research*, vol. 46, no. W1, pp. W11–W16, 2018.
- M. Imakaev, G. Fudenberg, R. P. McCord, N. Naumova, A. Goloborodko, B. R. Lajoie, J. Dekker, and L. A. Mirny, "Iterative correction of hi-c data reveals hallmarks of chromosome organization," *Nature methods*, vol. 9, no. 10, p. 999, 2012.