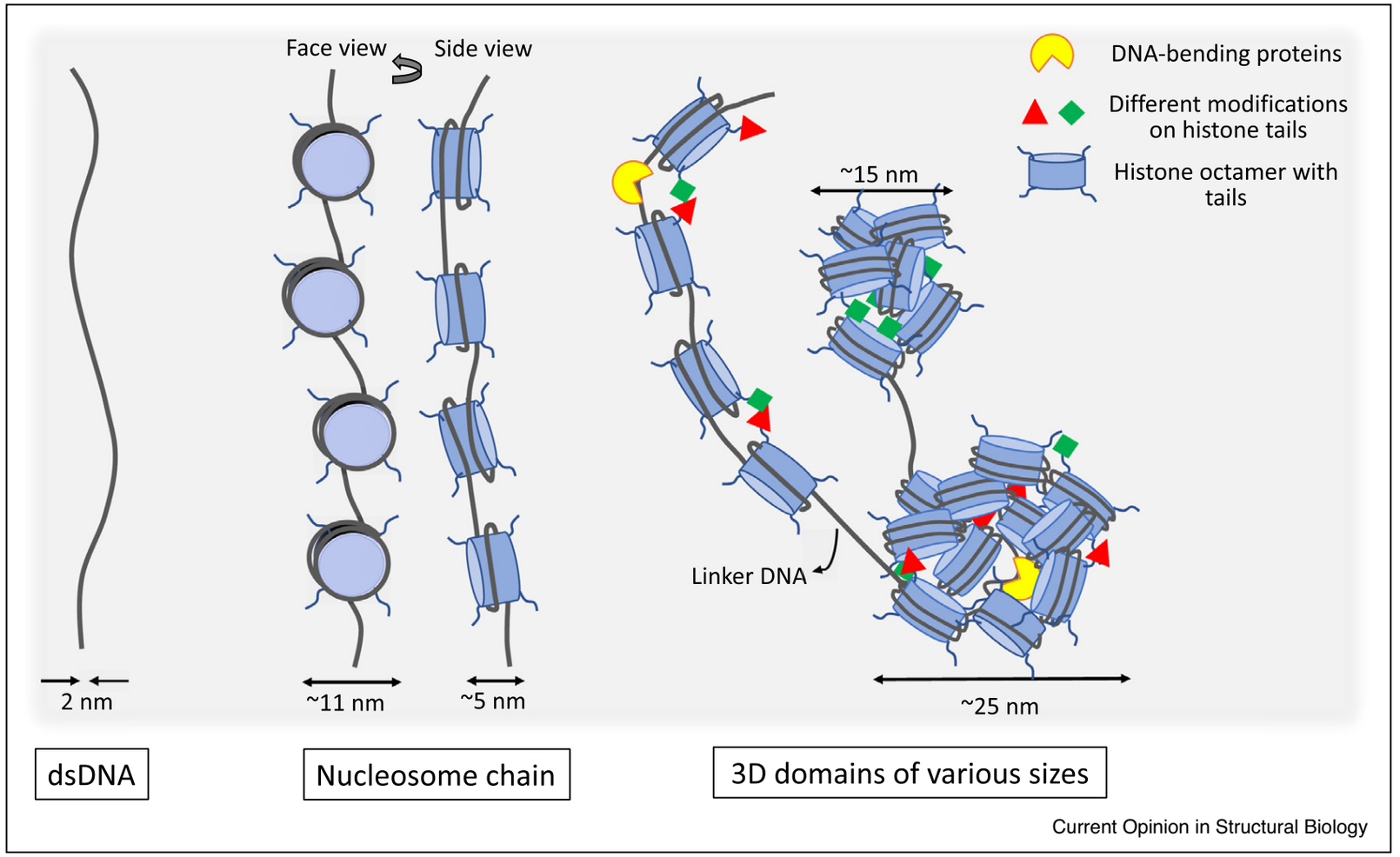
Physical models to understand chromatin assembly and inheritance of epigenetic information



Jyotsana J Parmar and Ranjith Padinhateeri

Current Opinion in Structural Biology 2020, **64**:111–118

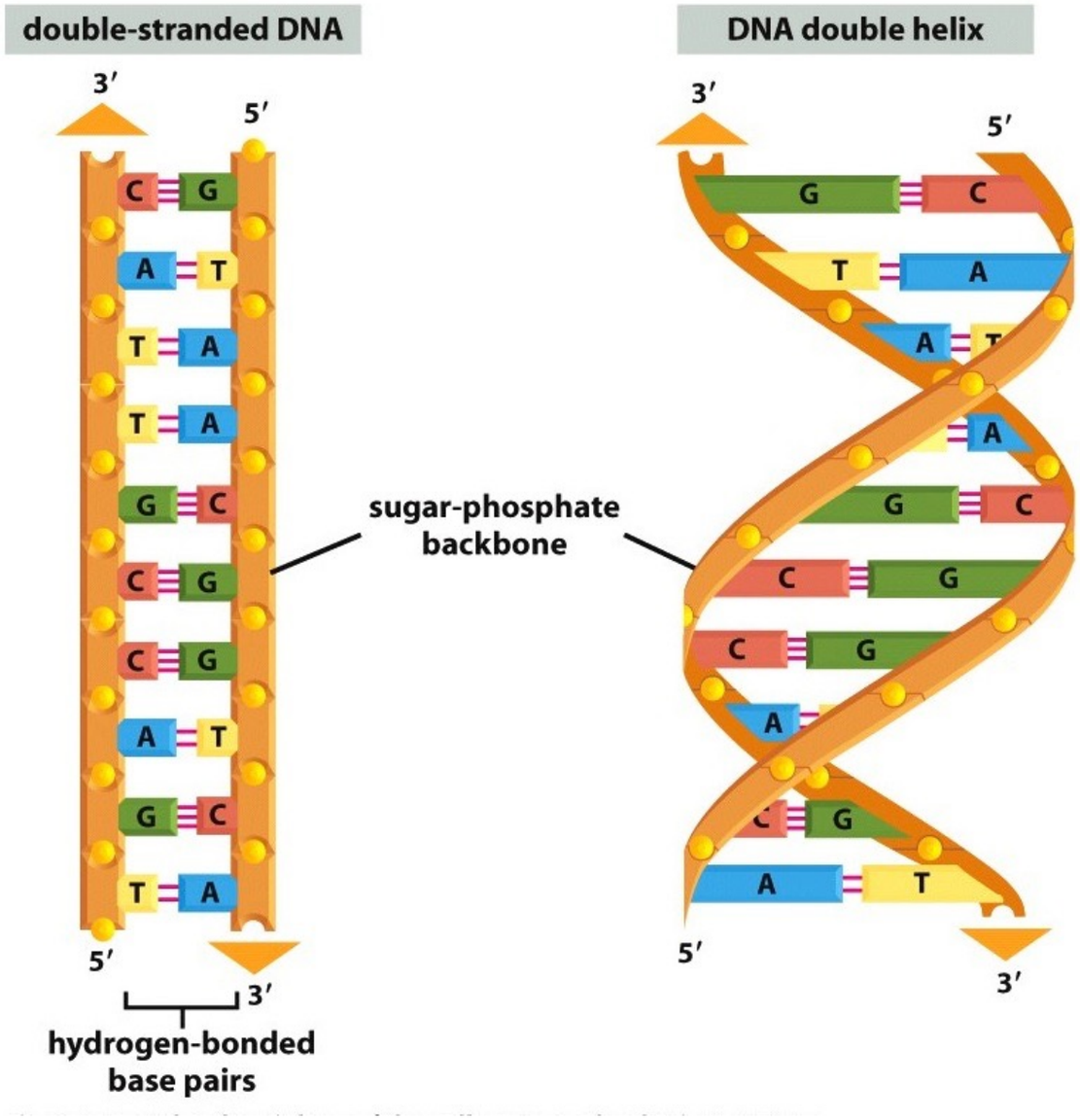
Ranjith Padinhateeri

Biosciences and Bioengineering

Indian Institute of Technology Bombay

www.bio.iitb.ac.in/~ranjith

DNA : molecule that contains code for cellular processes



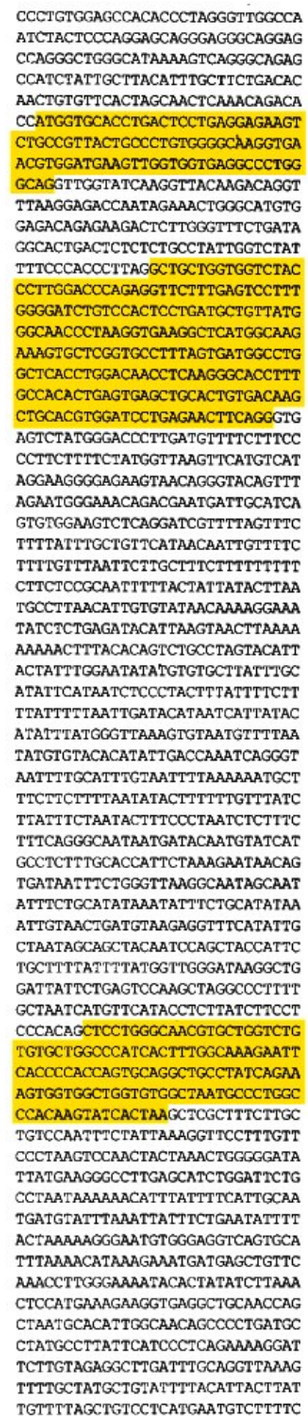
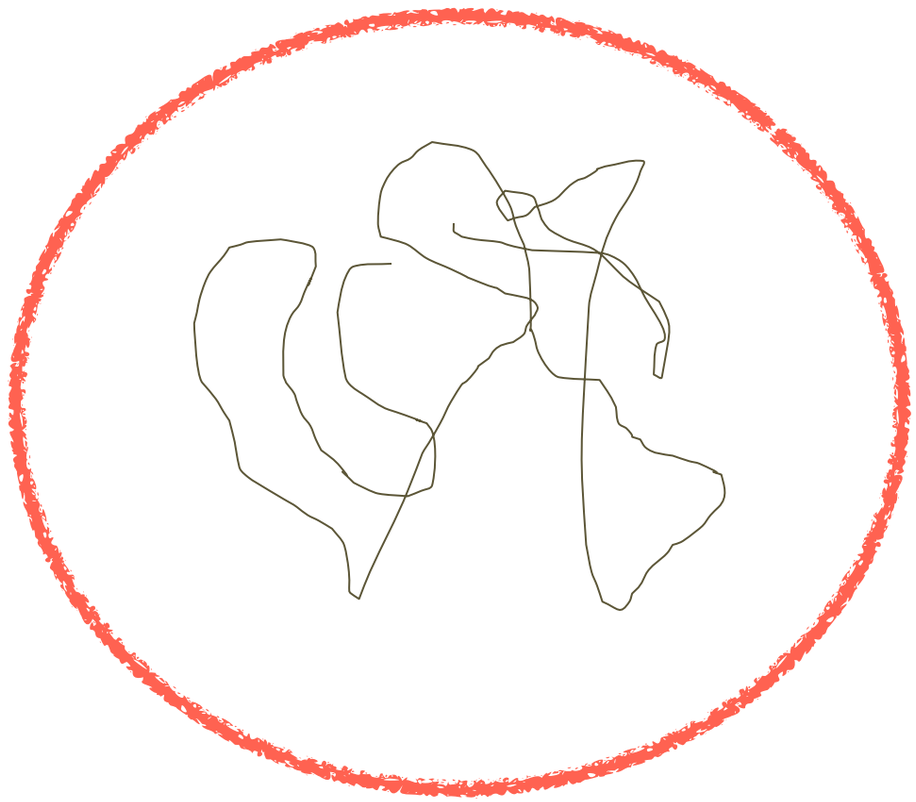
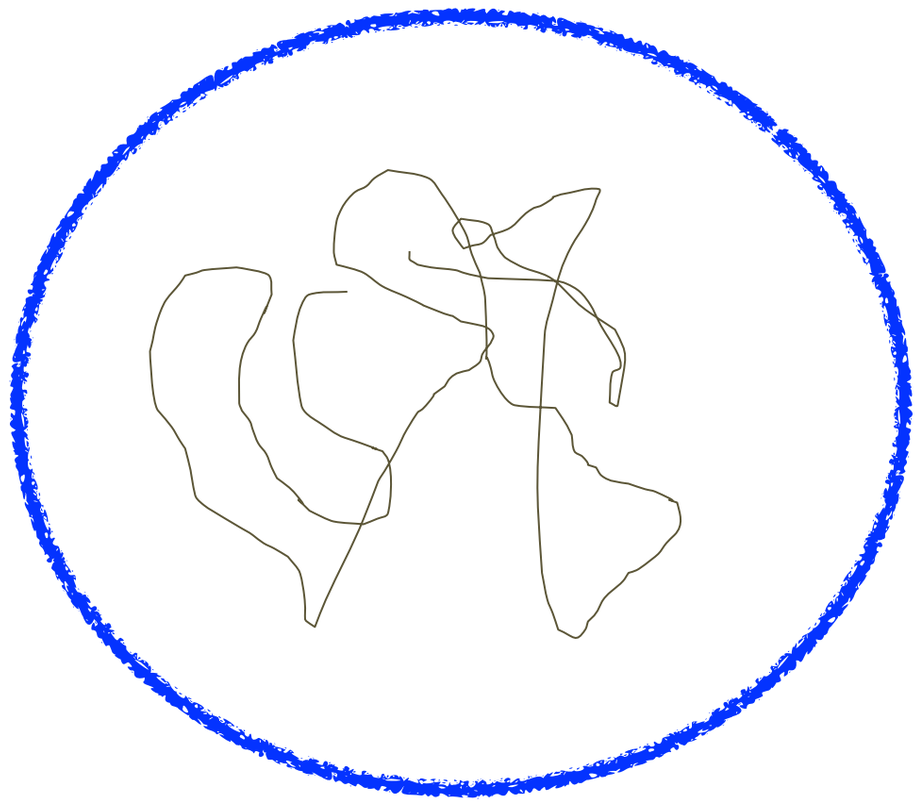


Figure credit: *Molecular Biology of the Cell* (© Garland Science 2008)

Different cell types; but same DNA



Cells in our skin

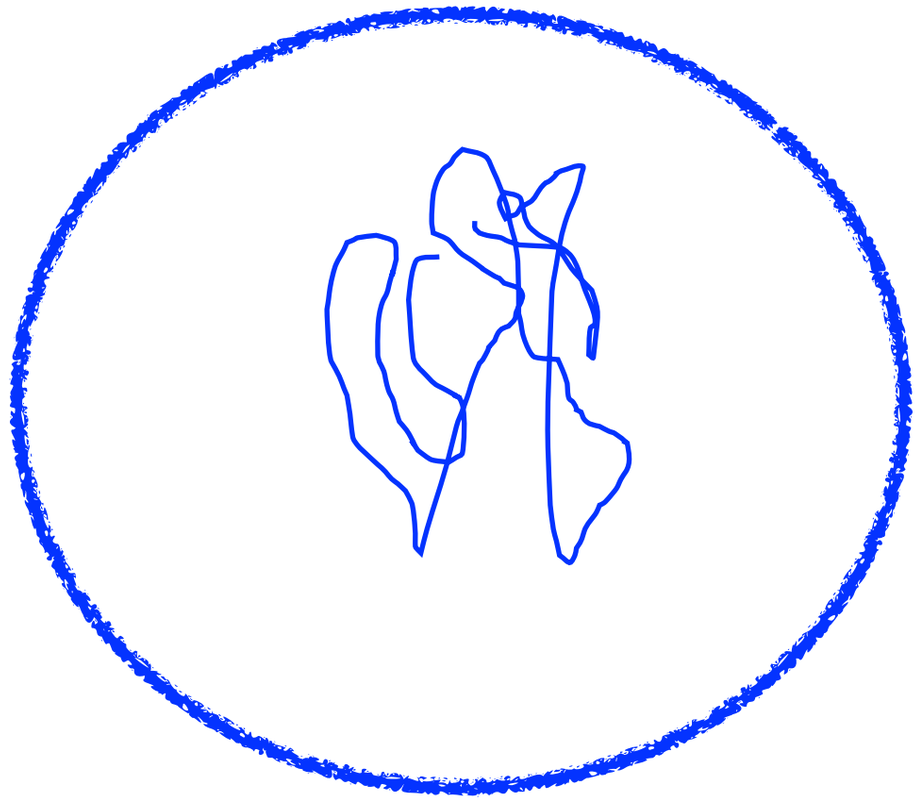


Cells in our eye

How do they show different behaviour ?

Chromatin is "assembled" differently in different cell types

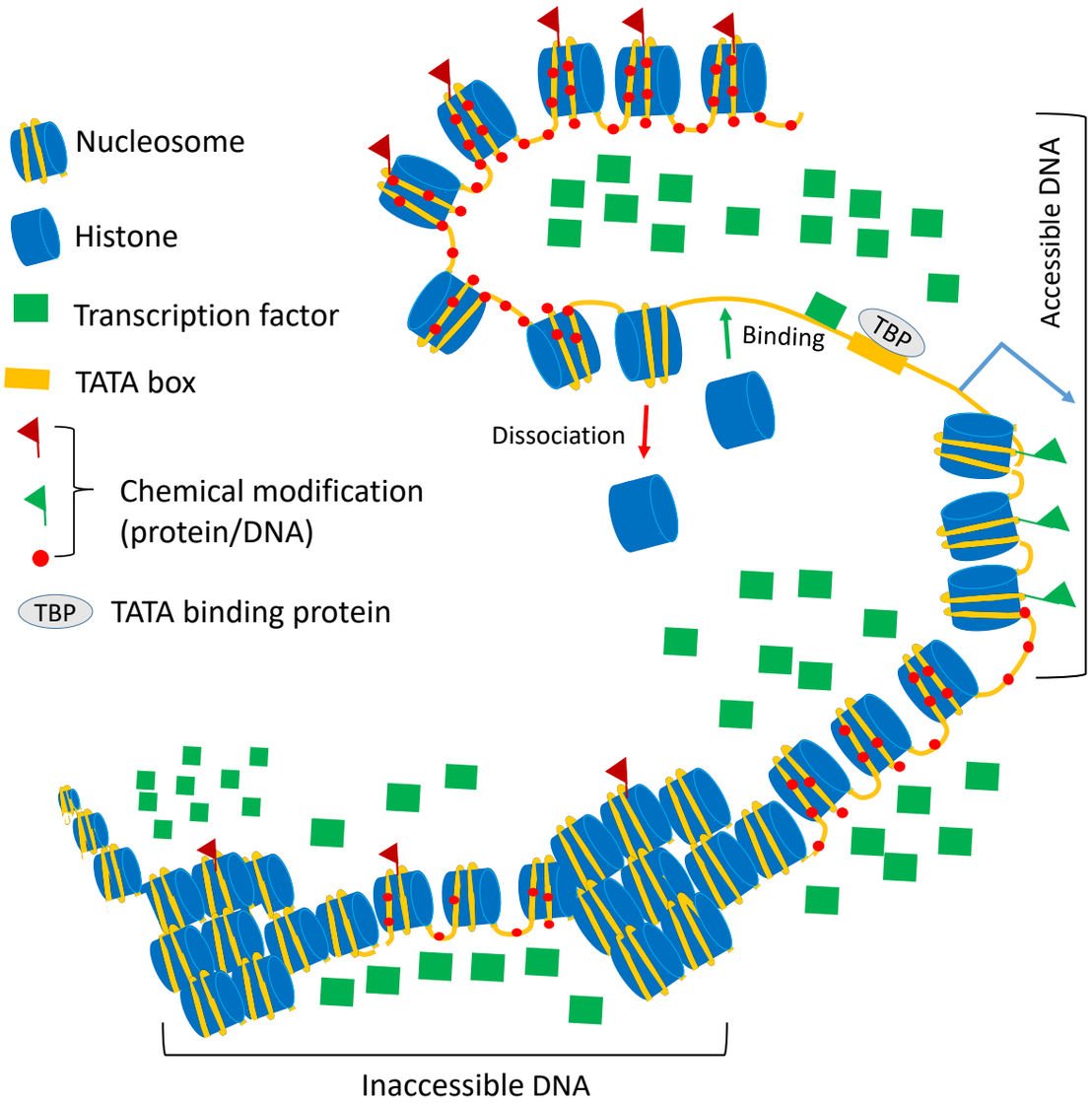
Cells in our skin



Cells in our eye

As a result, they express different sets of genes

Chromatin: Long polymer organization with multiple levels of information encoded



- Chromatin = DNA + protein
- Long polymer with heterogeneous interactions
- Different cell types (skin, brain) have different chromatin organization
- Different microstate and macrostates

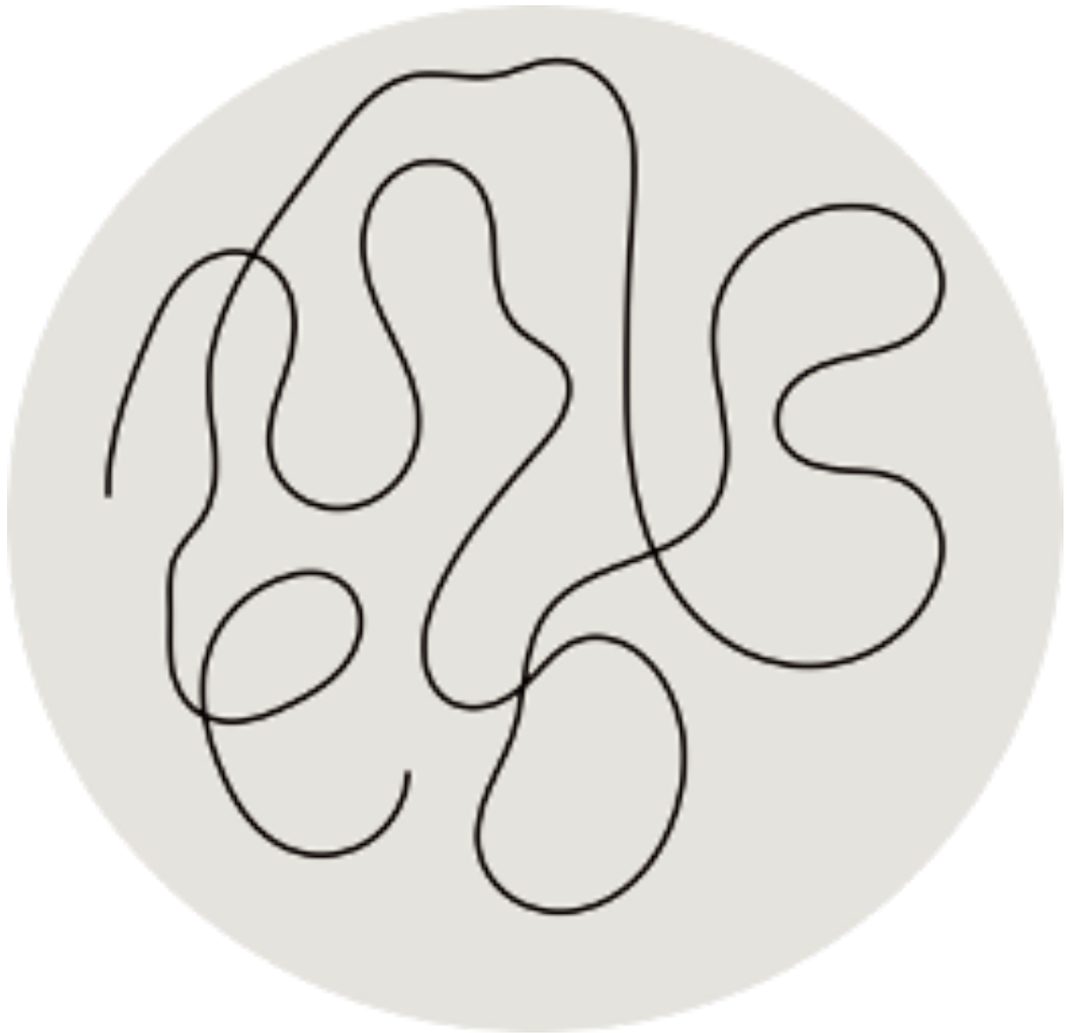
How is chromatin organised in 3D?

How chemical marks are organised along genome

From the known experimental data,

- What can we learn about the 3D organization of chromatin
- What can we say about copying epigenetic information before cell division

Experimentally measuring 3D organisation of chromatin

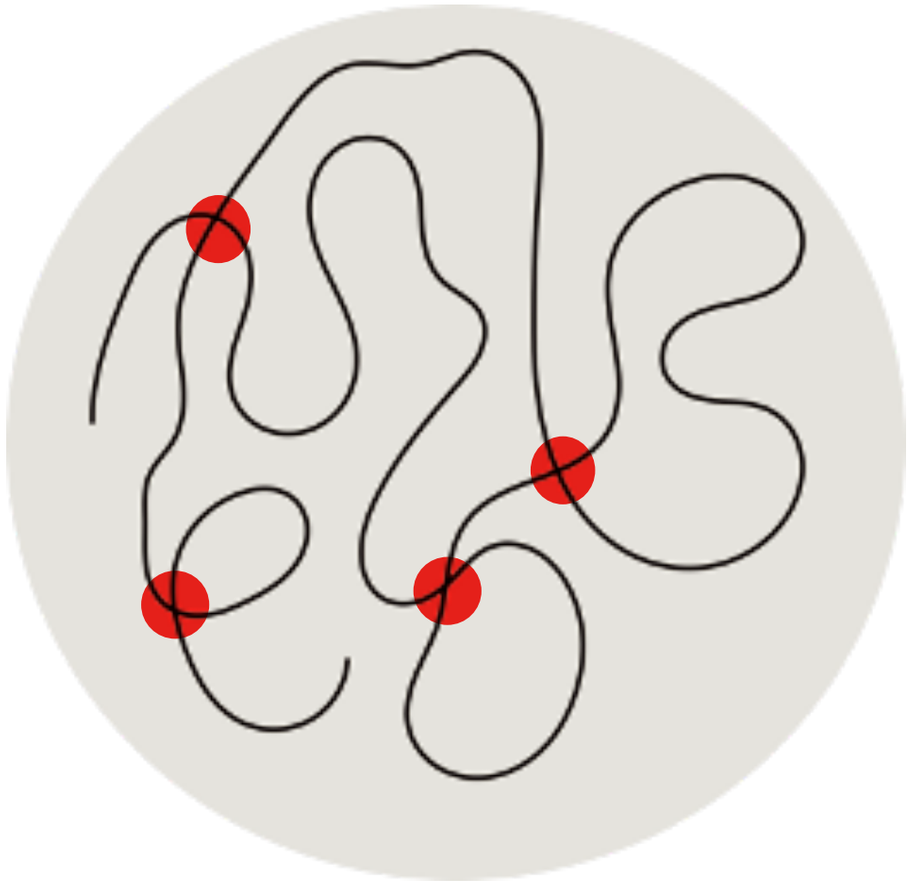


Lengthscale here: hundreds of kilo bases to mb

Chromatin conformation capture experiments

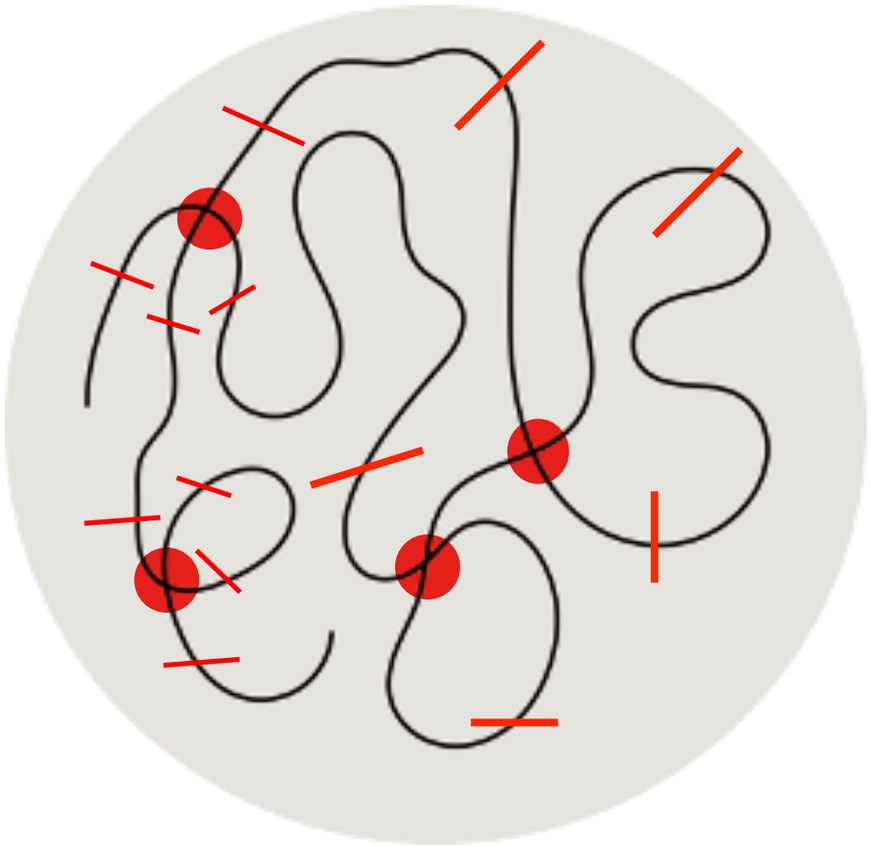
- Experiments can quantify the number of “contacts” between any two regions

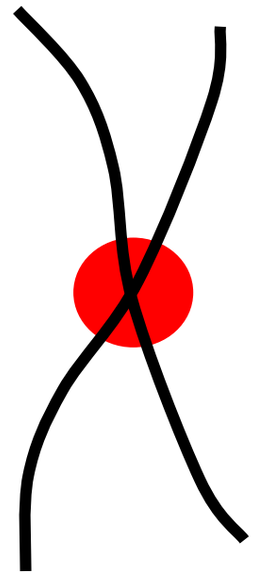
- Chromatin is cross linked (formaldehyde) at the locations of contact



Dekker et al, (2002) Science
Lieberman-Aiden et al, (2009) Science

Measuring contact probability between any two segments

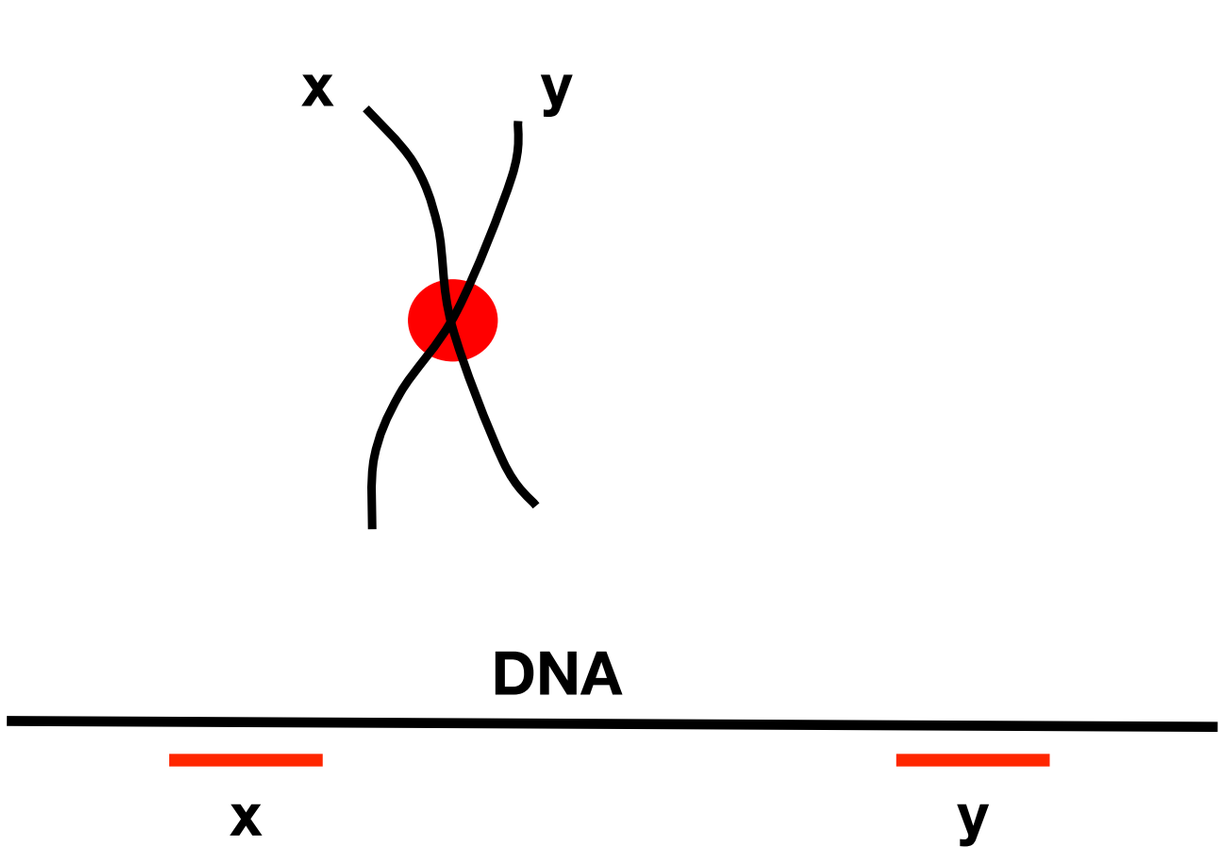


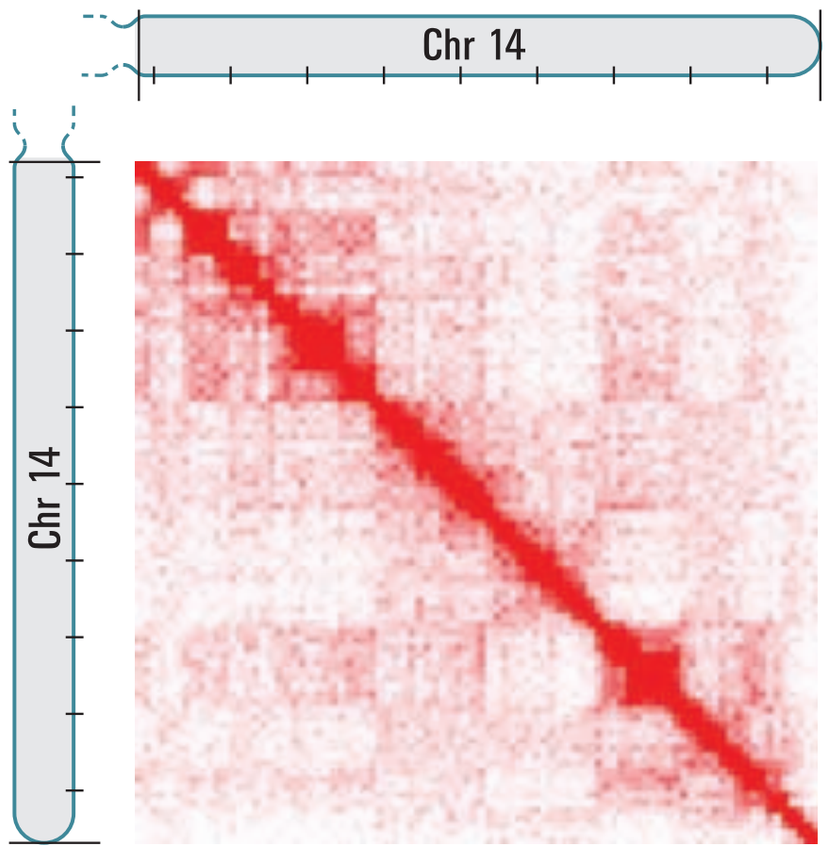


Cut the DNA using enzymes and separate cross-linked pieces; Sequence them

Data from a population of cells

Probability of contact between any two segments (x, y)



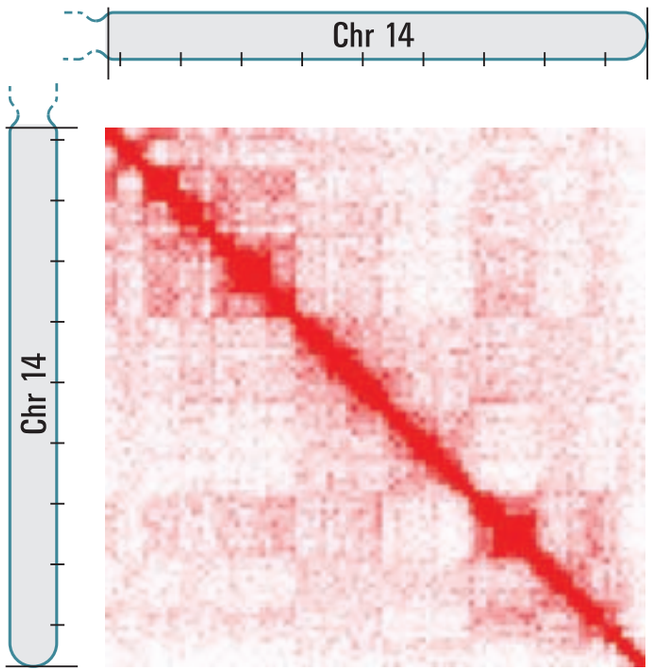


Lieberman-Aiden et al, (2009) Science

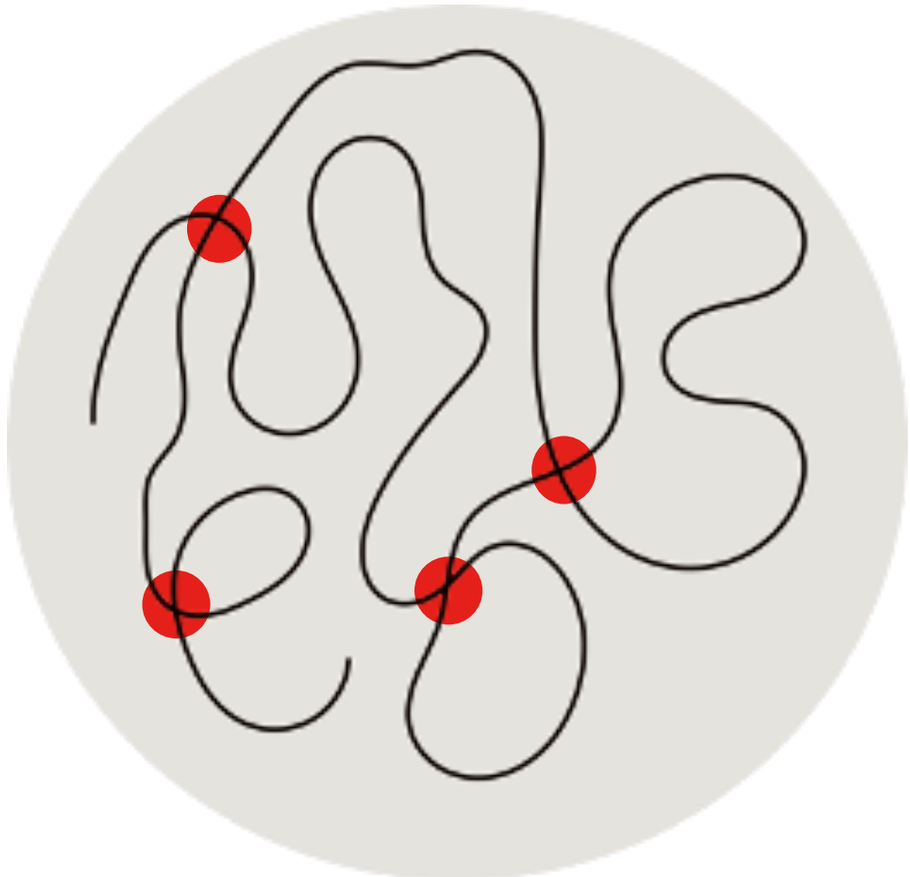
A symmetric matrix representing contact probability P(x,y)=P(y,x)

Data from a population of cells

How do we get 3D configuration of chromatin from a contact matrix?



?

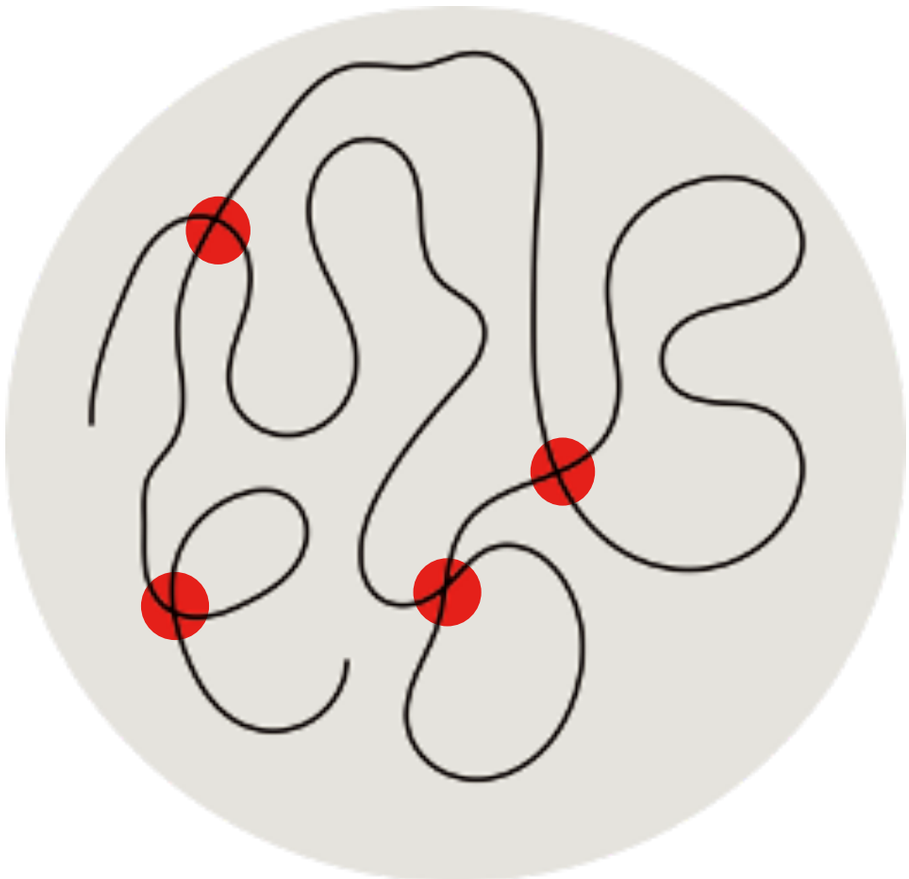


$p < 0.1$

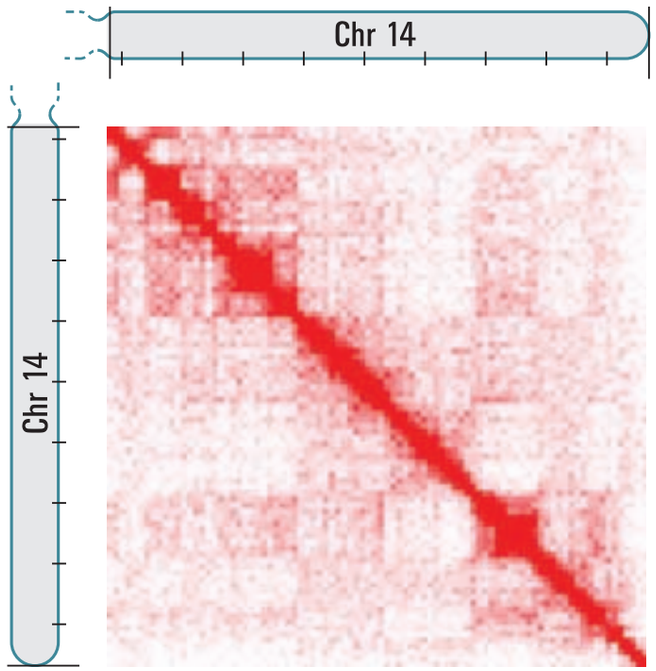
What is the 3D distance between two segments of my choice?

An “inverse” problem

Given a polymer with all interaction potentials, we can compute contact probability : “Forward problem”



Monte Carlo/
Brownian Dynamics



?

Known only the contact probability, we need to find interaction strengths between different segments such that the experimentally observed contact probability constraints are satisfied

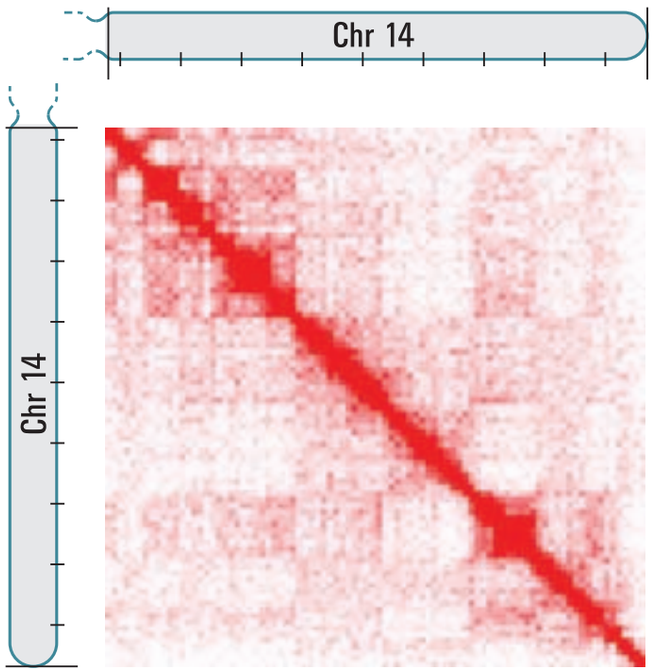
An Inverse Brownian Dynamics simulation to compute 3D organisation, given contact matrix

Kiran Kumari
IIT Bombay-Monash
Joint PhD program

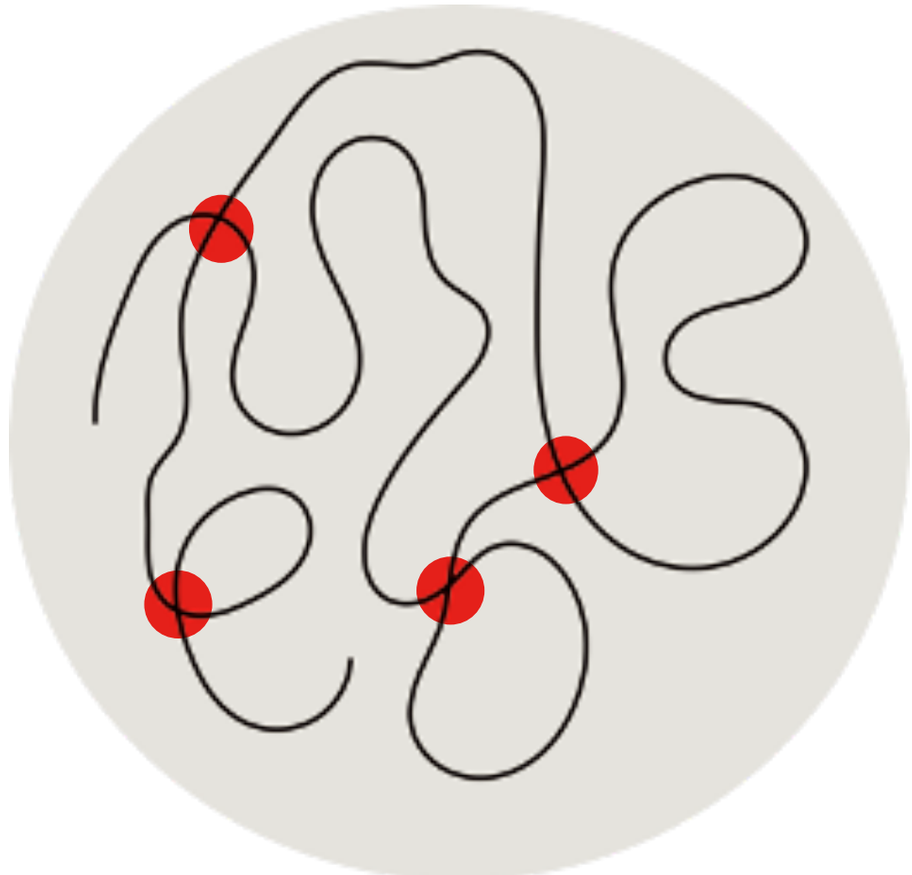
J. Ravi Prakash
Monash University

With
Burkhard Duenweg
MPI PR Mainz

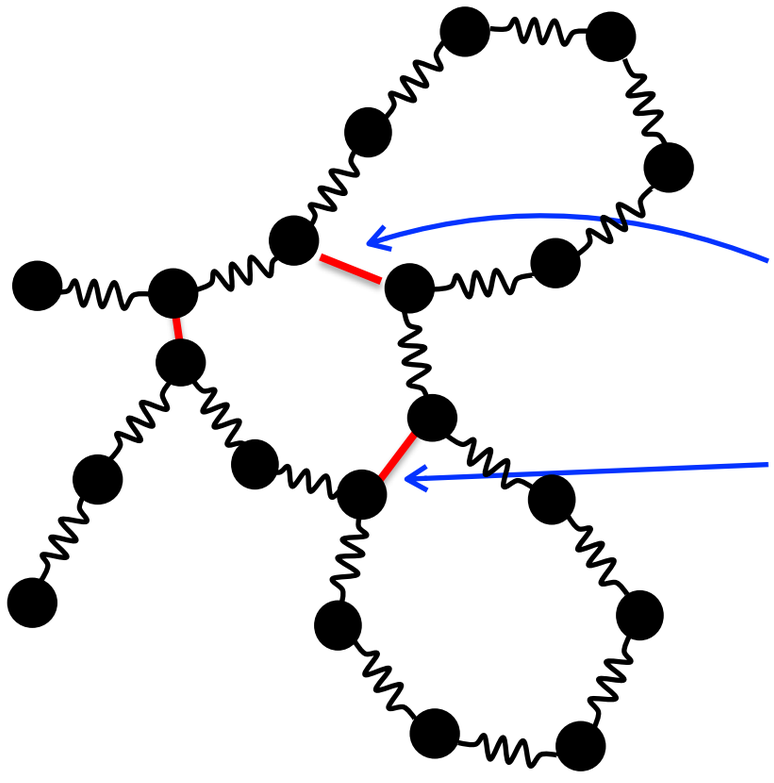
Aim: find out 3D configurations that satisfy this reference contact probability from experiments



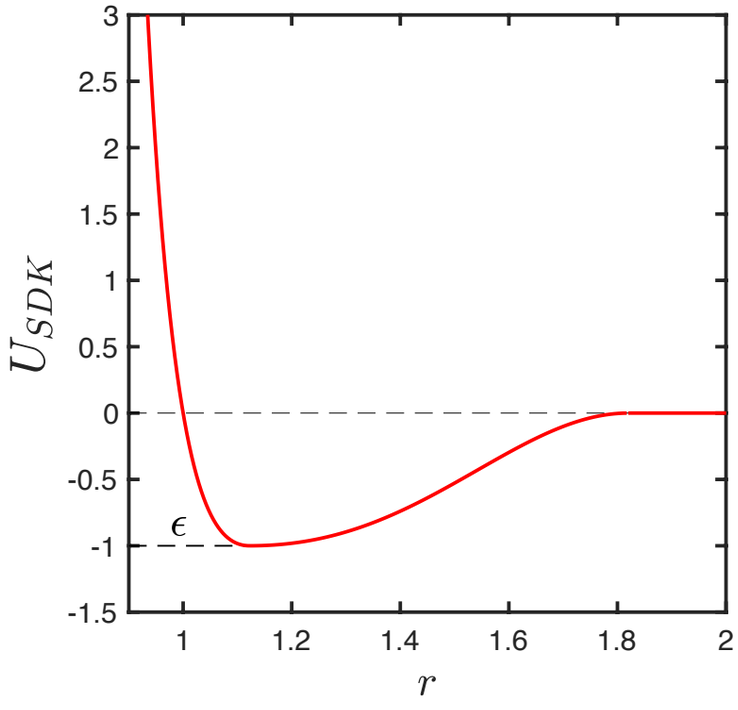
?



Chromatin as a bead-spring chain with attractive interactions between specific beads



Attractive interaction with a cut-off distance; representing intra-chromosomal interactions



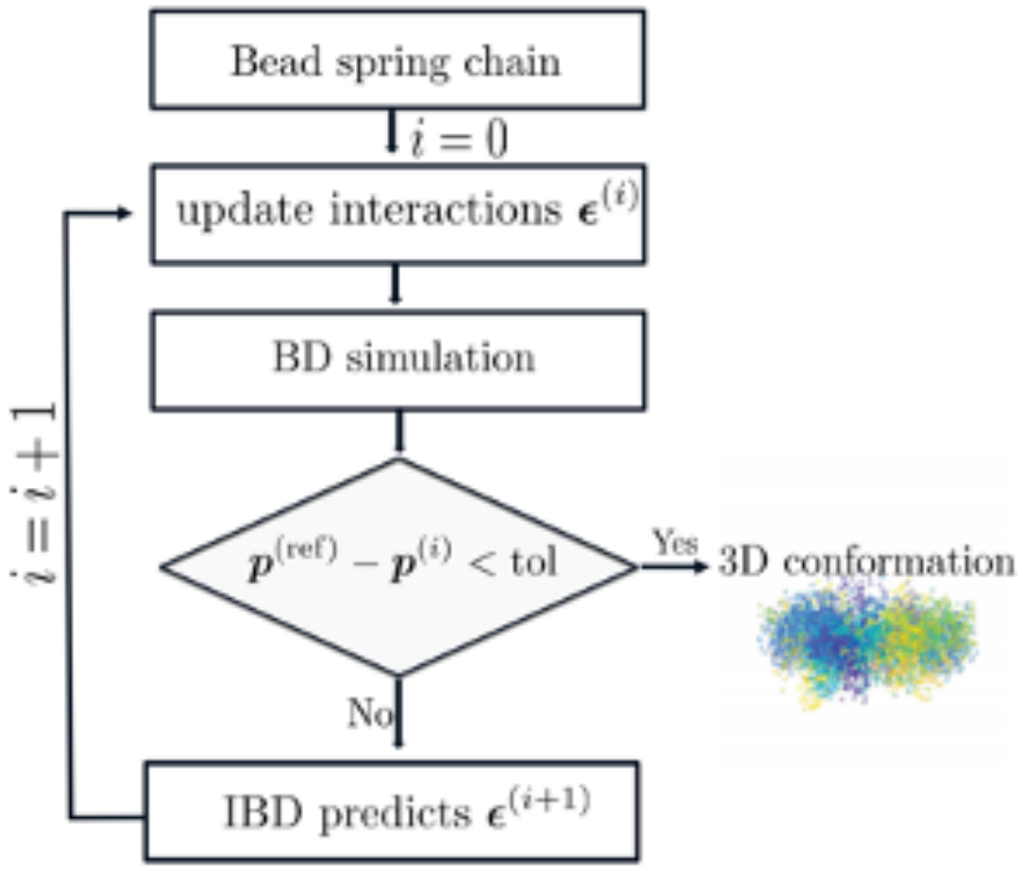
Beads: Excluded volume interactions

$\epsilon_{\mu\nu}$: Interaction strength between beads μ and ν

What is the optimal $\epsilon_{\mu\nu}$ such that we get back observed contact matrix?

Inverse Brownian Dynamics (IBD) algorithm

- Start with random interaction strength values ϵ
- Compute contact probability (forward simulation)
- Compare it with experimental value
- Tune interaction strengths until the contact probability is comparable to what is seen experimentally



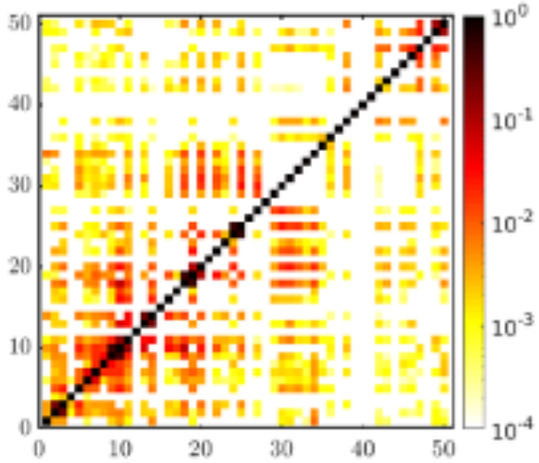
Optimal interaction energies that satisfy the experimentally known constraints

What is the 3D organization of the alpha globin gene locus?

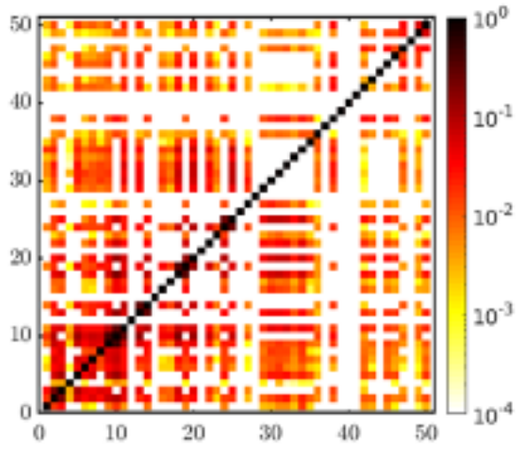
- The gene region that codes for Hemoglobin subunit alpha1

- 500kb region on human chromosome-16 (Encode region ENm008).

- Chromatin conformation capture experiments by Bau et al, Nat. Struct. Mol. Biol. (2011)



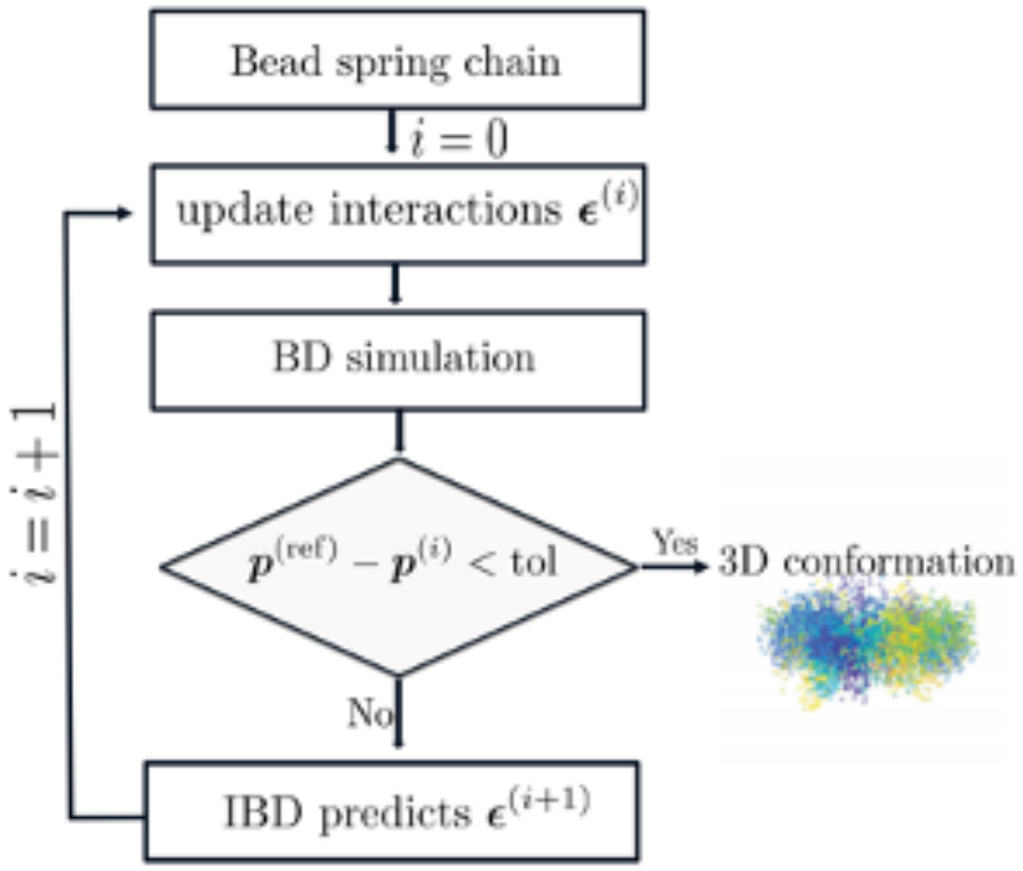
From K562 cells
Gene is "ON"
(Being read; proteins are being made)



From GM12878 cells
Gene is "OFF"
(Not being read; proteins are not made)

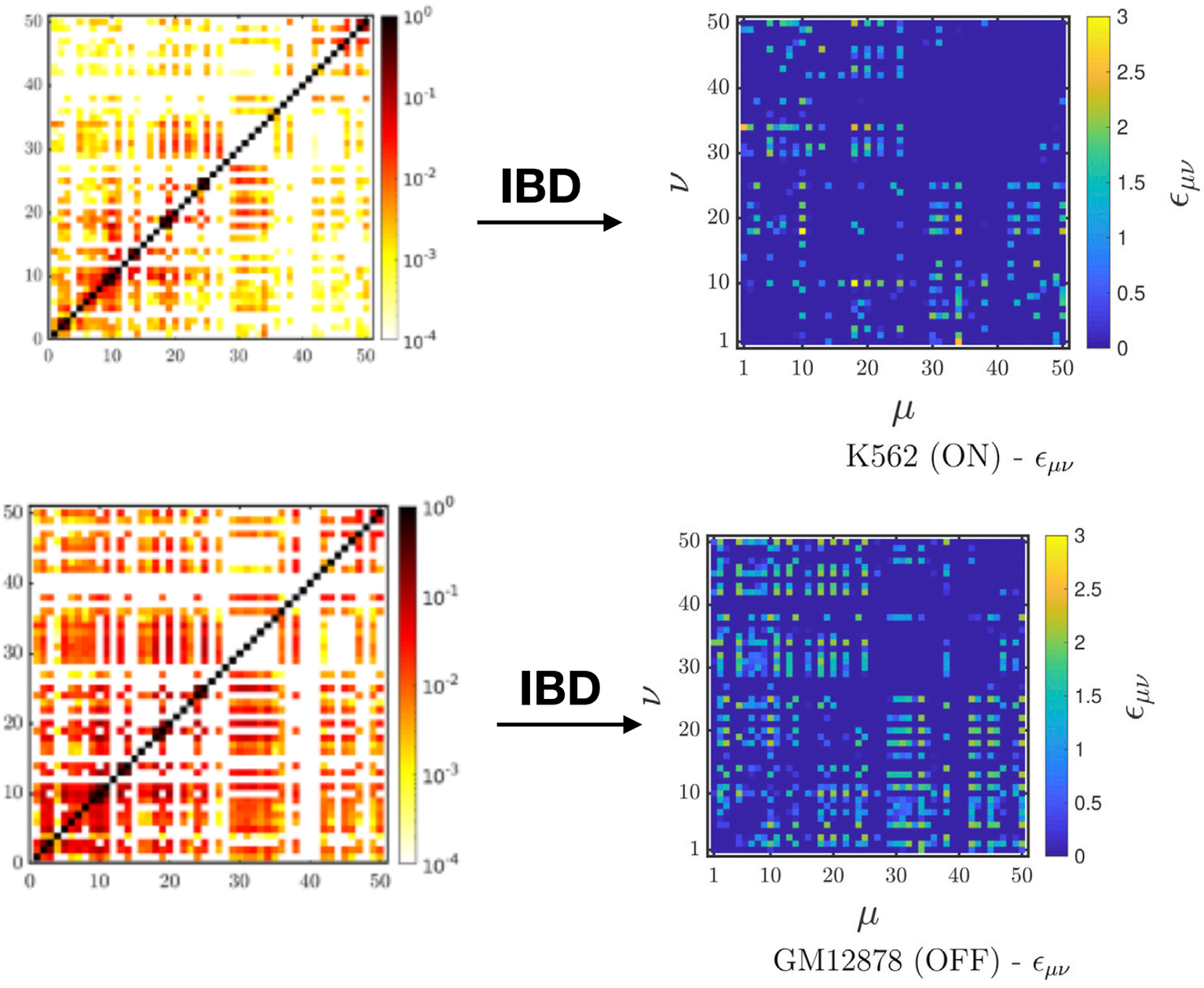
Inverse Brownian Dynamics (IBD) algorithm

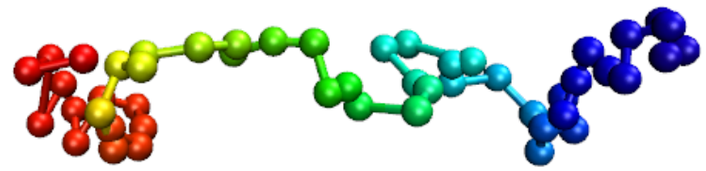
- Start with random interaction strength values ϵ
- Compute contact probability (forward simulation)
- Compare it with experimental value
- Tune interaction strengths until the contact probability is comparable to what is seen experimentally

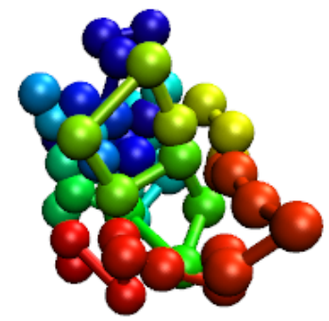


Optimal interaction energies that satisfy the experimentally known constraints

What is the 3D organization of the alpha globin gene locus?



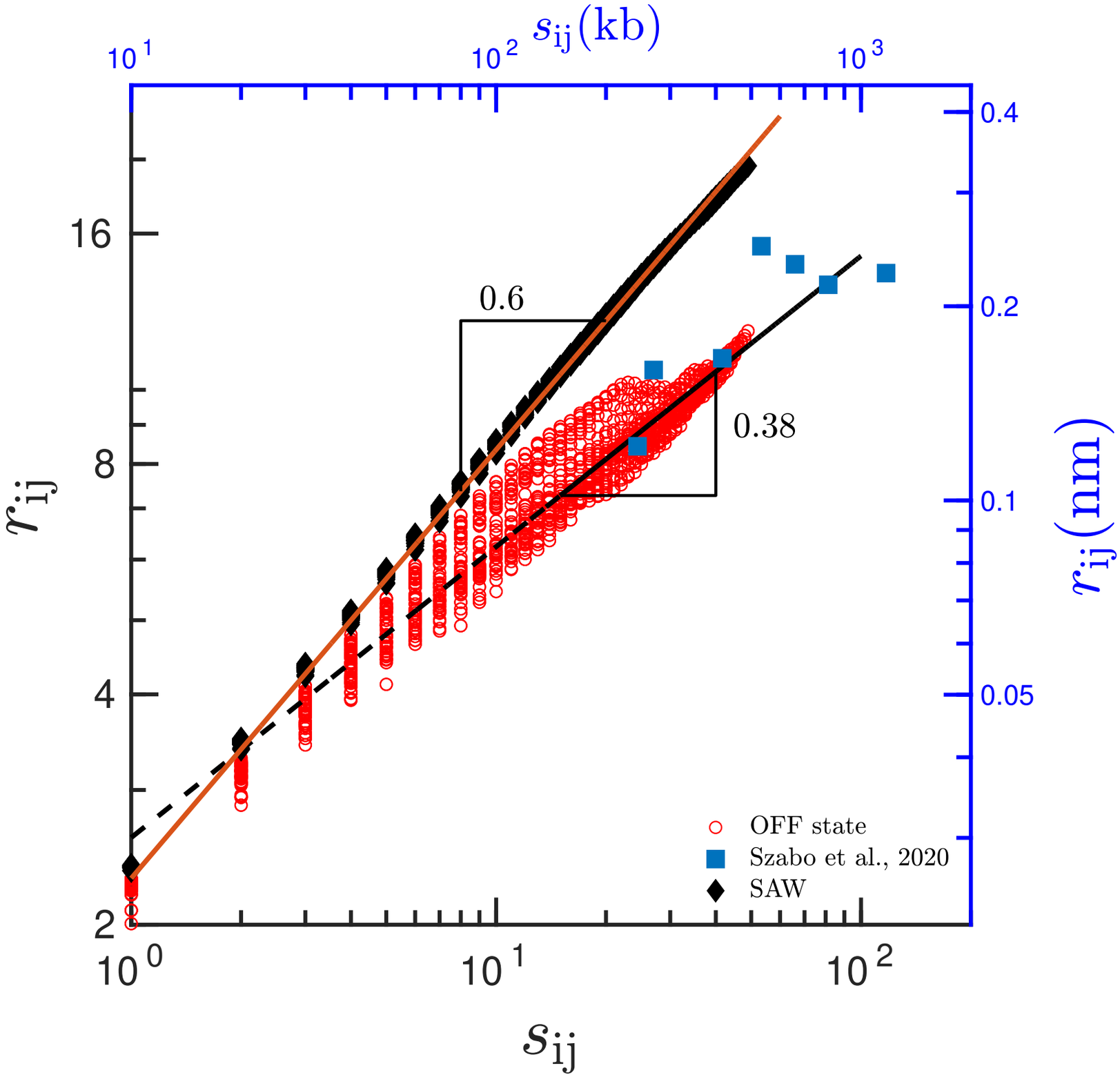




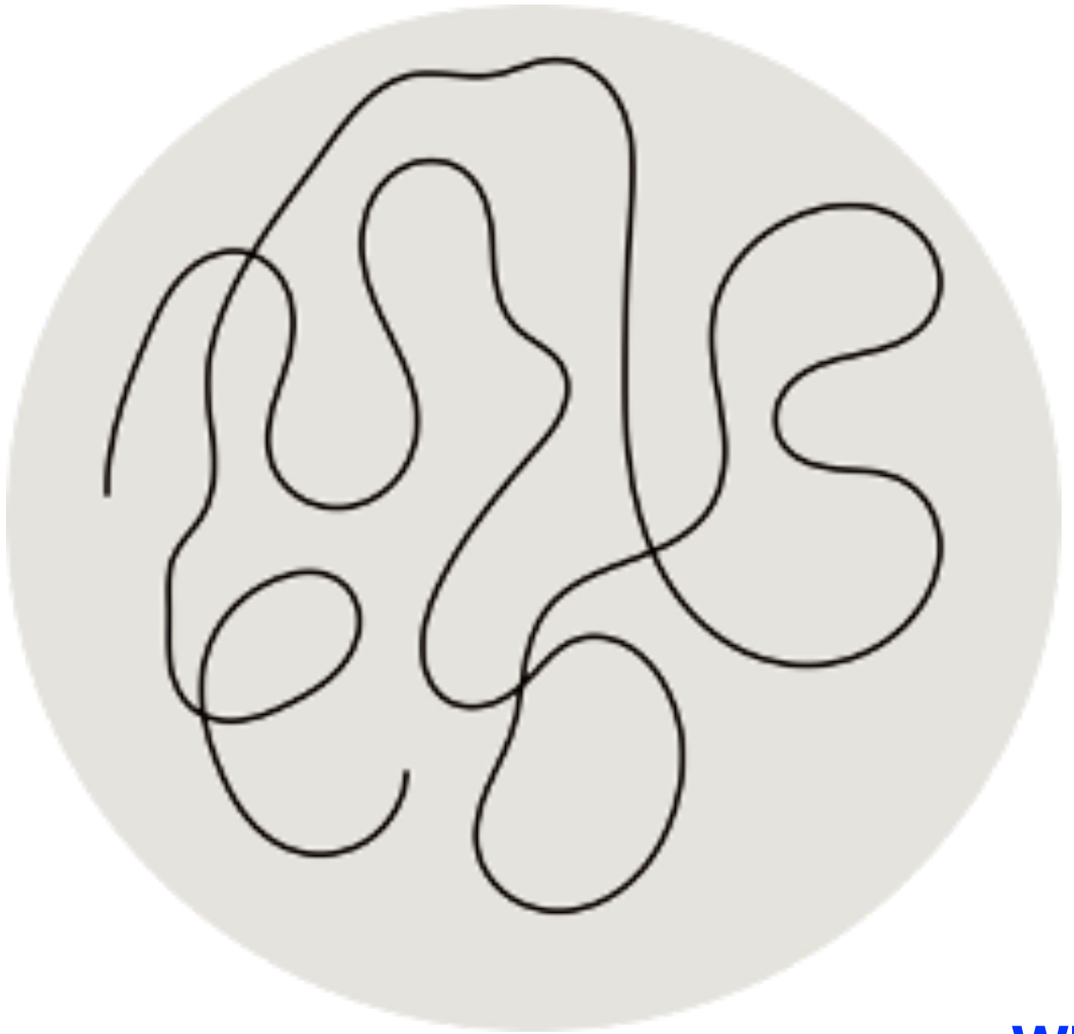
Typical snapshots

Kiran Kumari et al (2020) Biophys. J.

Scaling: Mean 3D distance vs genomic distance



Scaling relation between contact probability (P_c) and average 3D distance (R)



For an ideal chain

Looping probability $P_c(l) \sim l^{-\frac{3}{2}}$

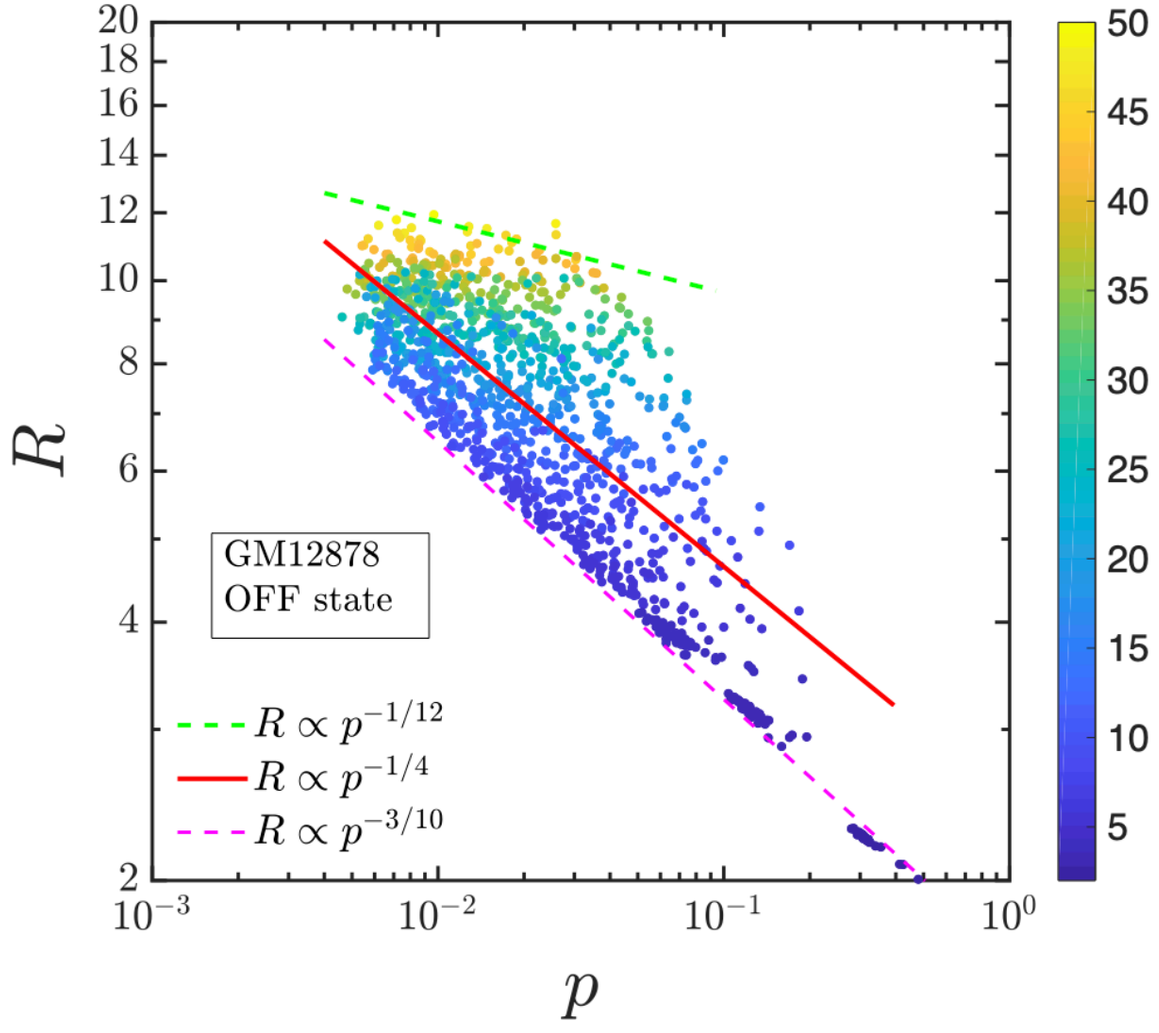
Average 3D distance $R^2 \sim l$

$$\Rightarrow P_c \sim R^{-3}$$

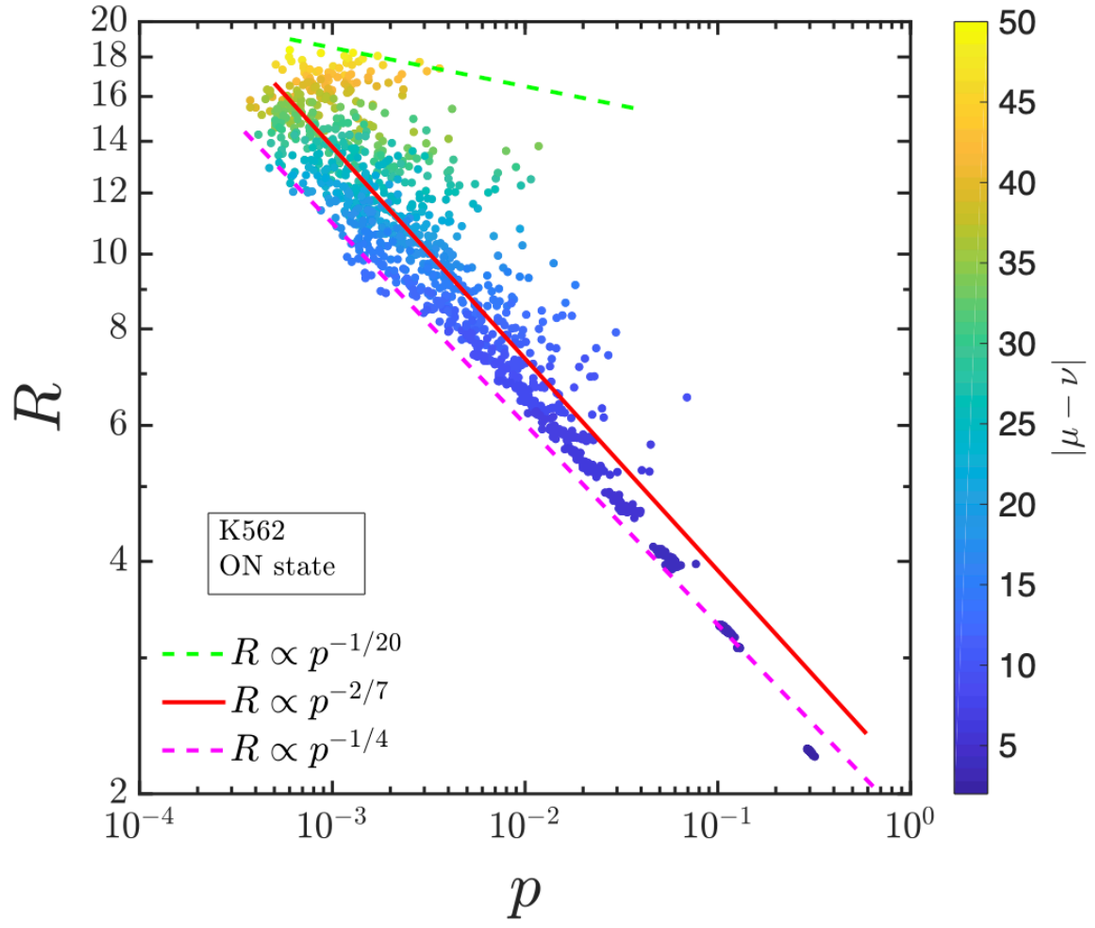
$$\Rightarrow R \sim P_c^{-\frac{1}{3}}$$

What is this relation for a chromatin segment?

Average 3D distance: function of contact probability, contour distance between segments (color), and interaction strengths



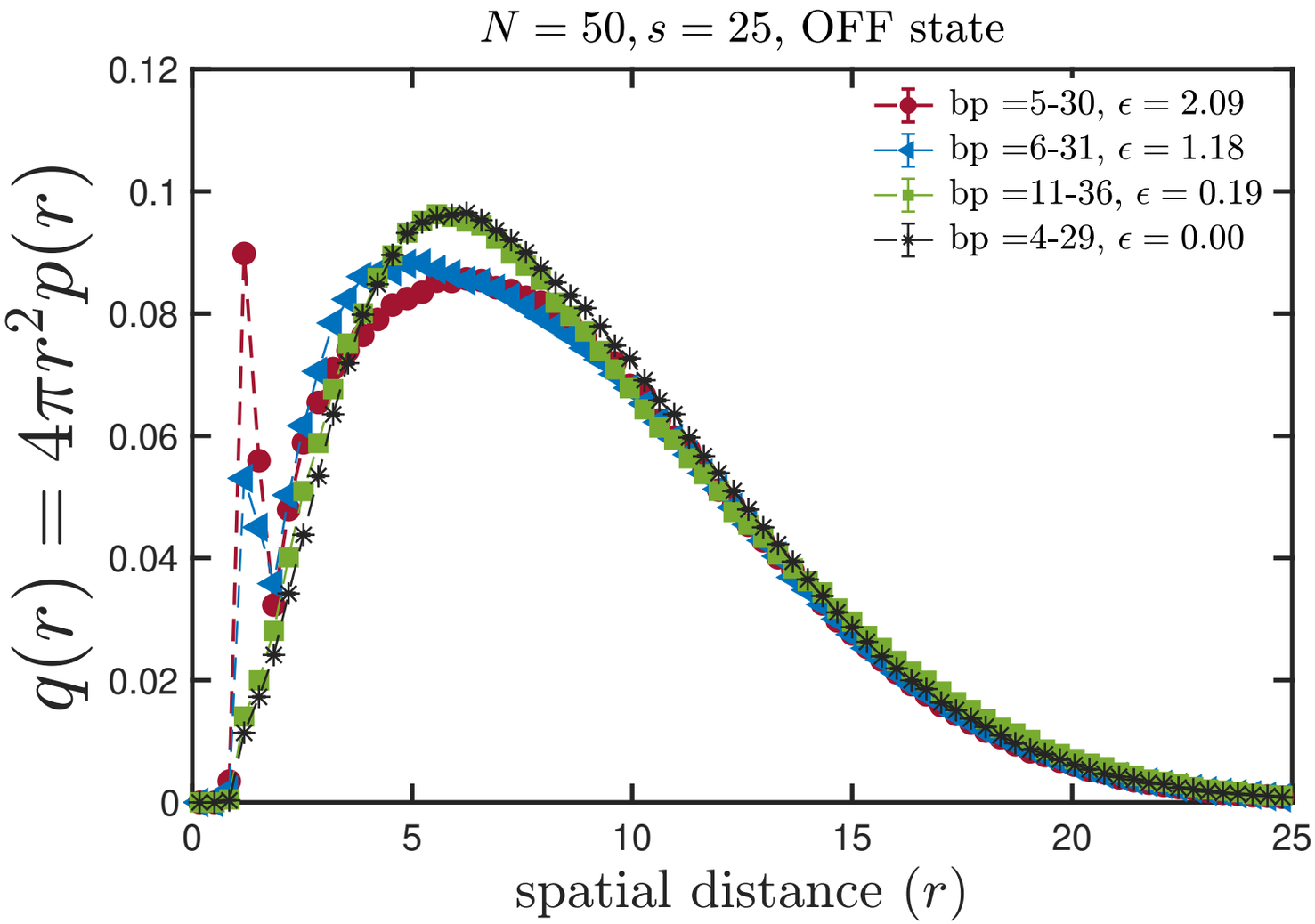
Variation around $R \sim P_c^{-\frac{1}{4}}$



Variation around $R \sim P_c^{-\frac{1}{3.5}}$

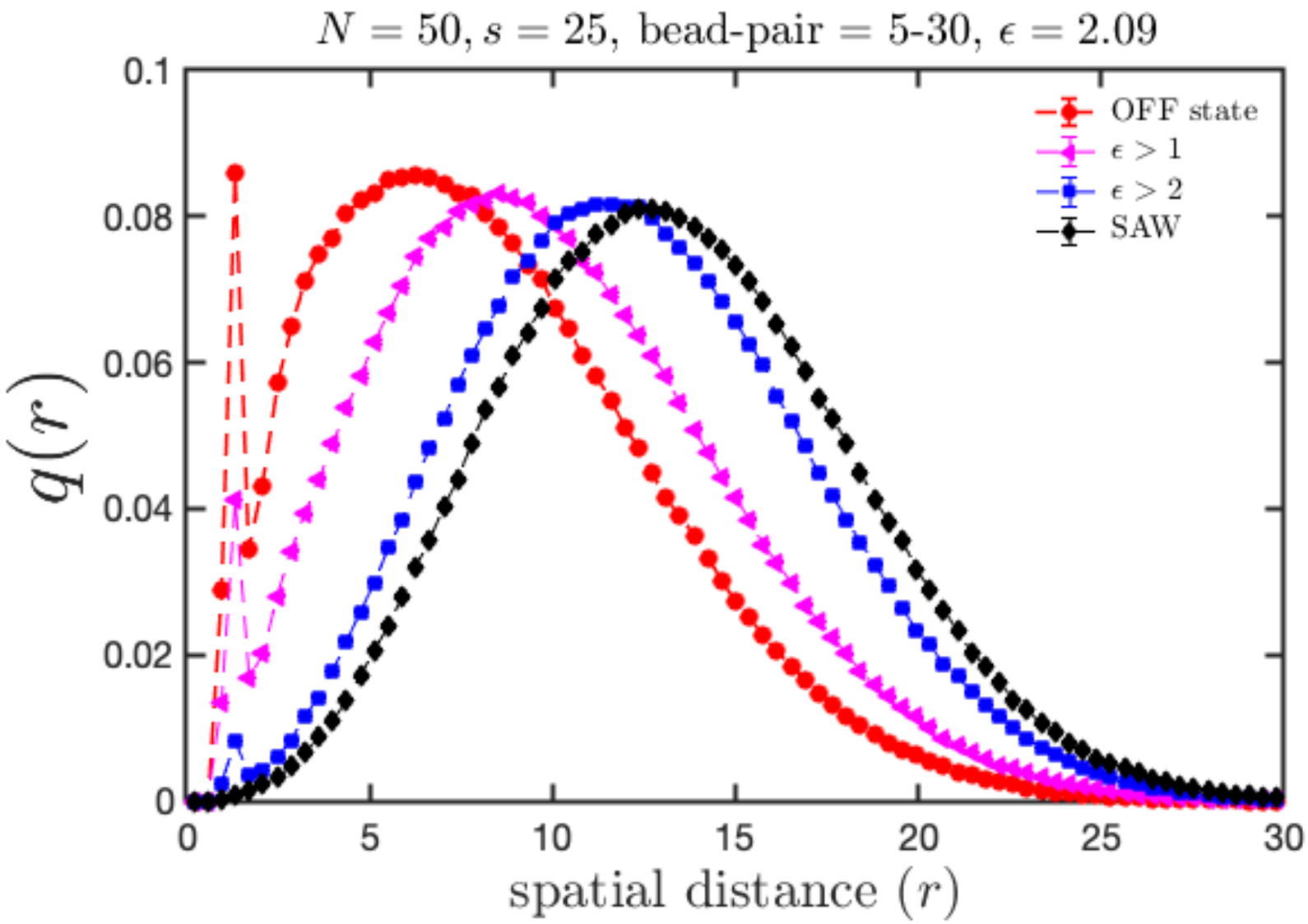
Kiran Kumari et al (2020) Biophys. J.

3D distance distribution between a specific pair of points



Average distances may not tell the full story!

Other interactions collectively determine the distance distribution between a pair of beads



In biology, proximity of far away genes is crucial; enhancer and promoter

Summary-I: Computing 3D organization from contact maps

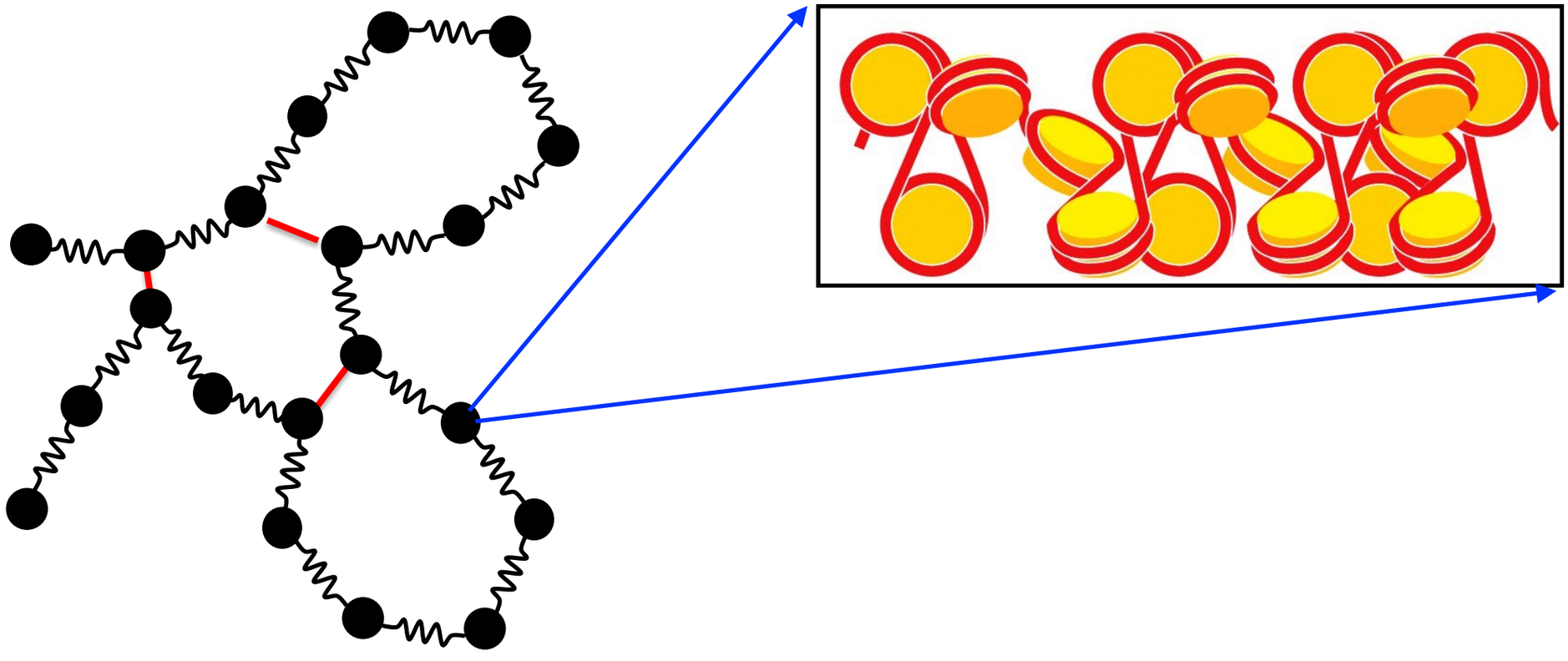
- Experiments measure contact probability between segments of chromatin

- Inverse problem: determining the interactions and 3D configurations such that the experimentally seen contact probabilities are satisfied

- Configurations of Alpha globin gene

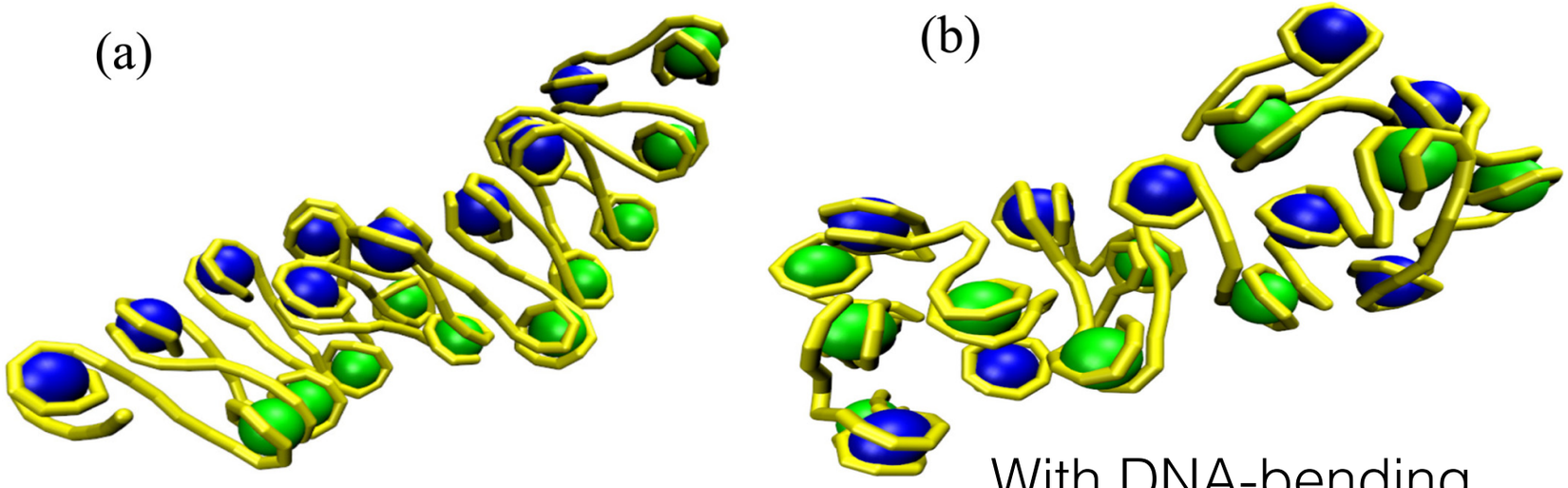
- Average 3D distance is a function of contact probability, contour distance between segments, and interaction strengths

If you zoom in, there is organization at a different scale



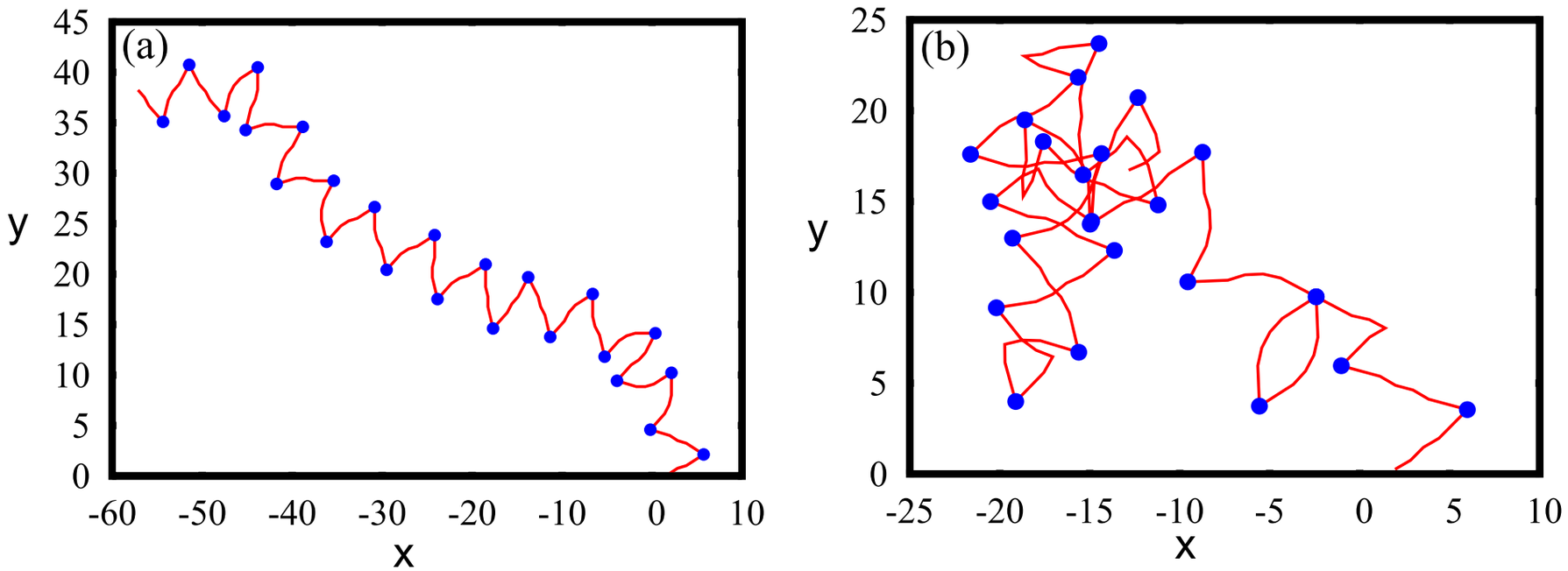
Picture Molecular Biology of the Cell 5/e (© Garland Science 2008)

Our earlier work: plausible explanation for absence of regular 30nm chromatin structure



Without any DNA-bending non-histone protein

With DNA-bending Non-histone protein



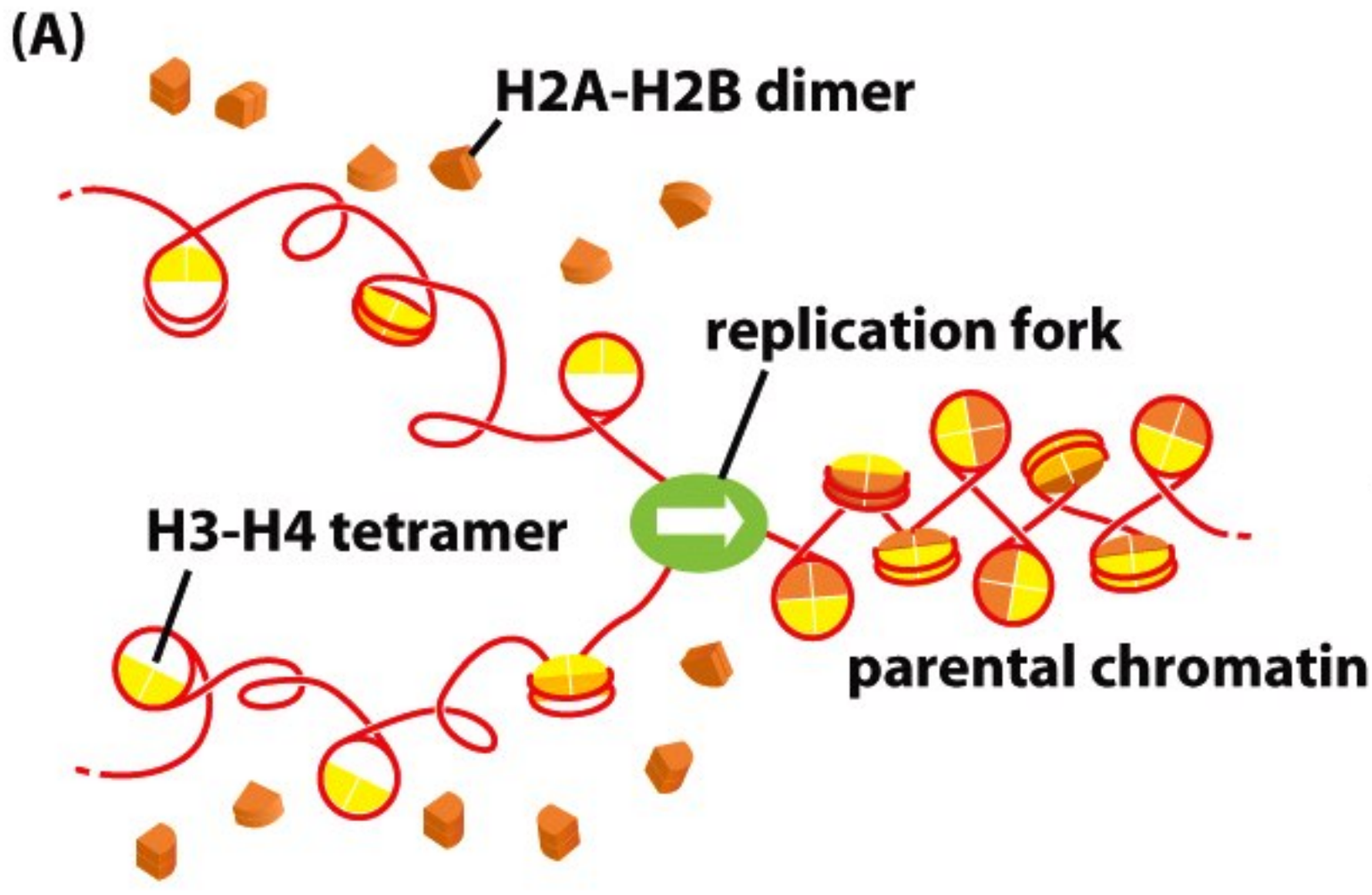
Bajpai, Jain, Inamdar, Das and Padinhateeri (2017) PLOS Comp. Bio.
Bajpai and Ranjith Padinhateeri (2020) Biophys. J

Part II: What can we say about copying chromatin information before cell division

When cells divide, DNA (genetic code) is copied.

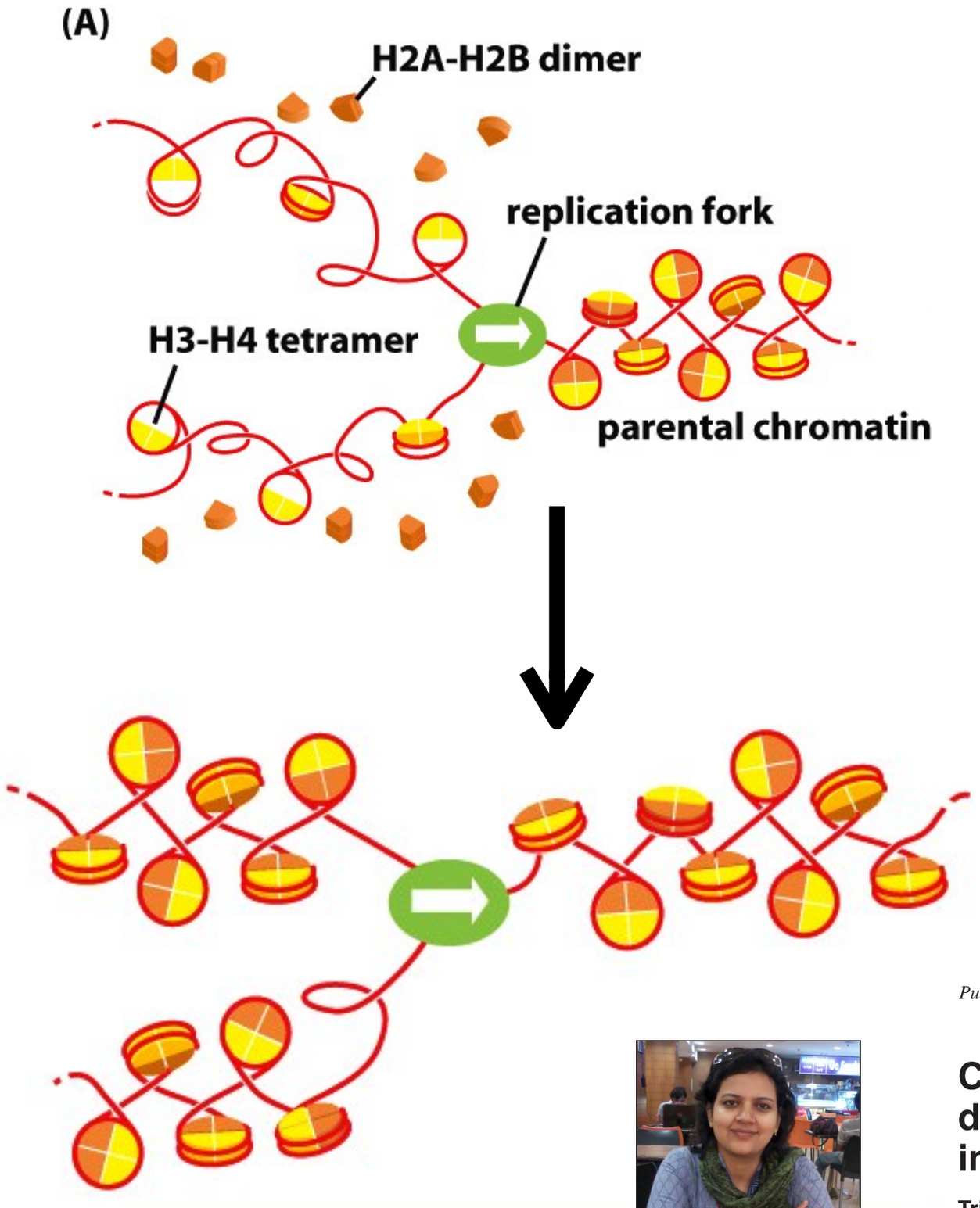
What happens to the epigenetic information?

Copying DNA before cell division (DNA replication)



Chromatin is disassembled! How do you assemble it back?

Re-assembling chromatin after replication



How do you assemble nucleosomes back at the right location?

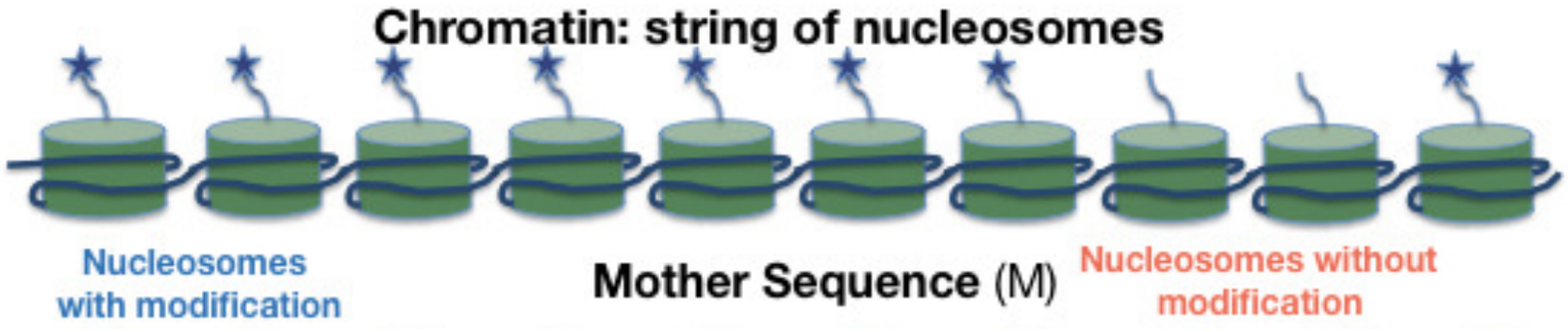
Published online 21 March 2018

Nucleic Acids Research, 2018, Vol. 46, No. 10 ***4991–5000***
doi: 10.1093/nar/gky207

Coupling of replisome movement with nucleosome dynamics can contribute to the parent–daughter information transfer

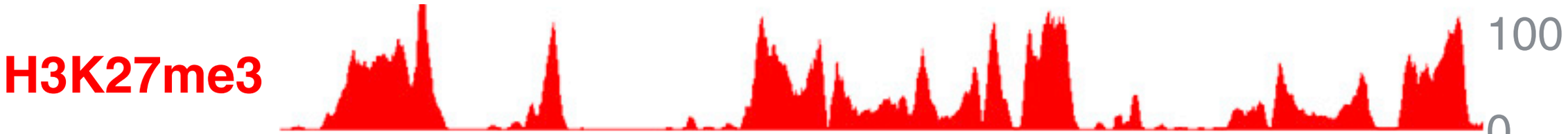
Tripti Bameta[1,*], Dibyendu Das[2] and Ranjith Padinhateeri[3,*]

Each nucleosome has chemical modifications like a “flag” (acetylation/methylation)

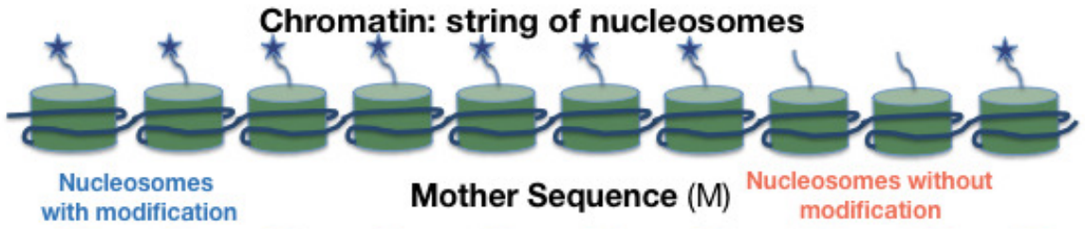


The sequence of of this modification (the pattern) encodes information on how to fold chromatin; e.g., it decides the local “interaction strength”

Typical histone modification pattern (population averaged)



Single cell: **1111100001110000001101110110100100100100010101010111**



Reverón-Gómez et al., 2018, Molecular Cell *72*, 239–249
October 18, 2018
https://doi.org/10.1016/j.molcel.2018.08.010

When cells divide, DNA (genetic code) is copied.

What happens to the epigenetic information?

- Certain modifications (many repressive marks) are “inherited” during cell division

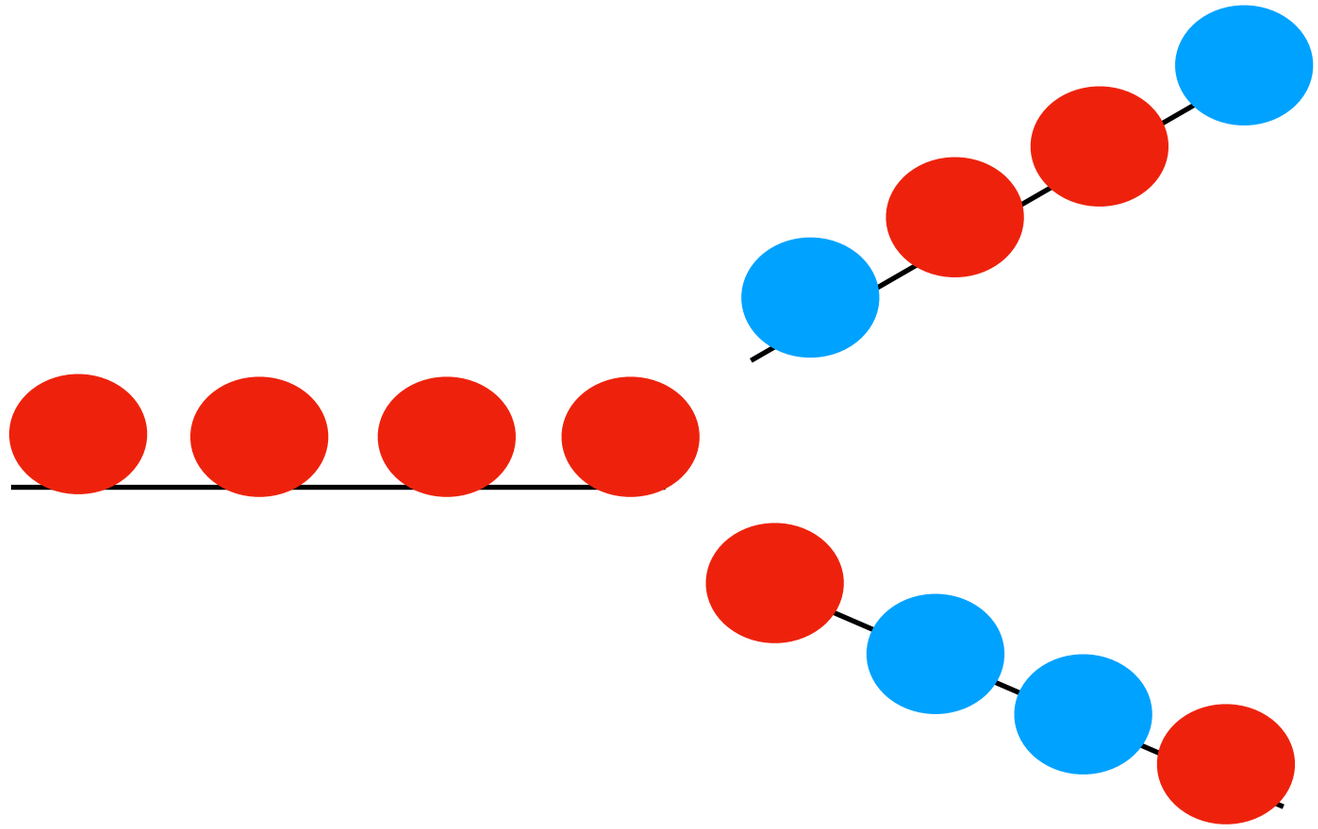
- Certain other marks are not inherited but re-established (somehow) after replication

Groth Lab, Mol Cell 2018,
Reinberg Lab, Science 2018, 2019

What happens to histone modifications during replication?

- During replication, each nucleosome from mother chromatin is randomly placed on one of the two daughter chromatin strand (probability 0.5).

- Half the nucleosomes are newly assembled

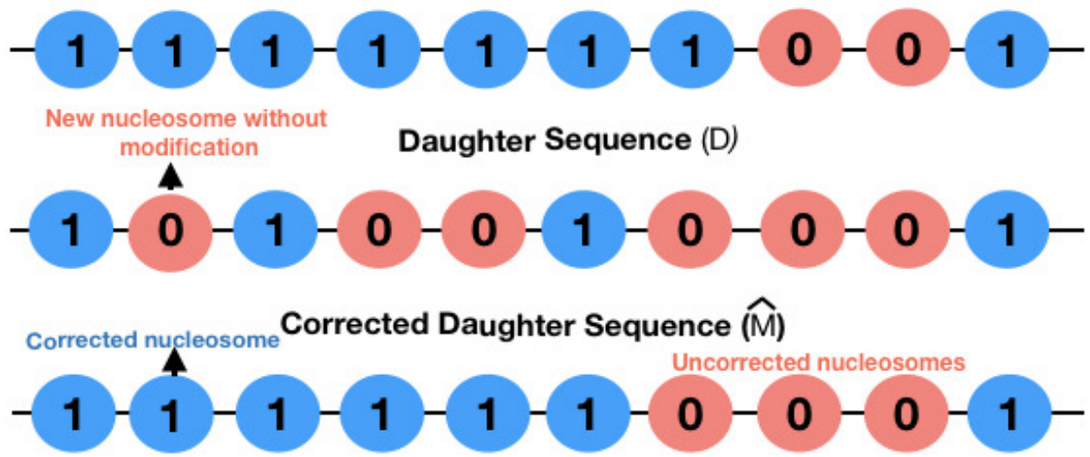


Ramachandran and Henikoff 2015

Groth lab, Science 2018

What happens to histone modifications during replication?

- During replication, each nucleosome from mother chromatin is randomly placed on one of the two daughter chromatin strand (probability 0.5).



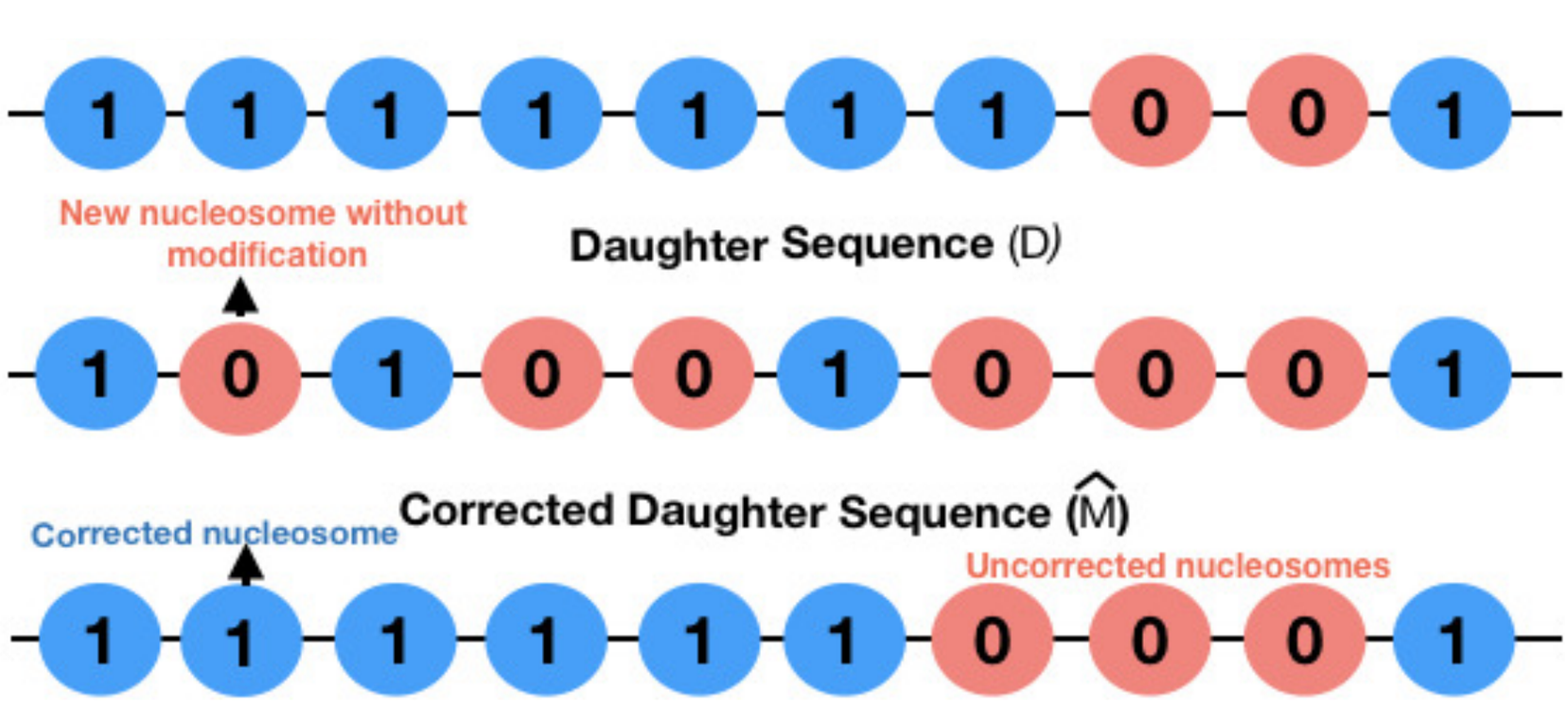
- Half the nucleosomes are newly assembled
- Old nucleosome will carry modification (1) if it has any.
- New nucleosome will NOT have any modification (0).

$$\mathbf{D} = \mathbf{M} \cdot \mathbf{Z}$$

AND operation with
Z=Random (IID) binary sequence

Groth lab, Science 2018

Given a daughter-chromatin histone modification sequence D, how can a cell reconstruct mother-like modification pattern ?

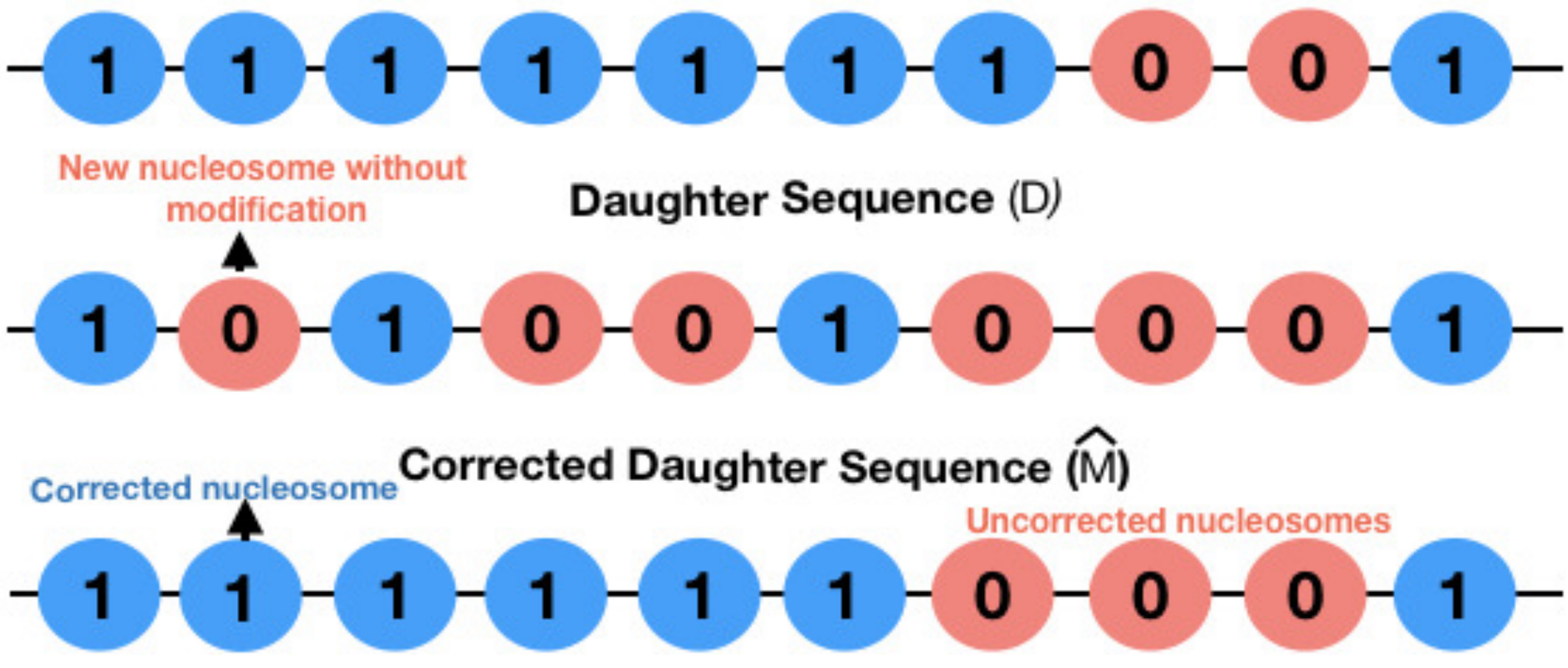


Reconstructing back such sequences, known noisy version of the sequence, is a problem in communication/information theory

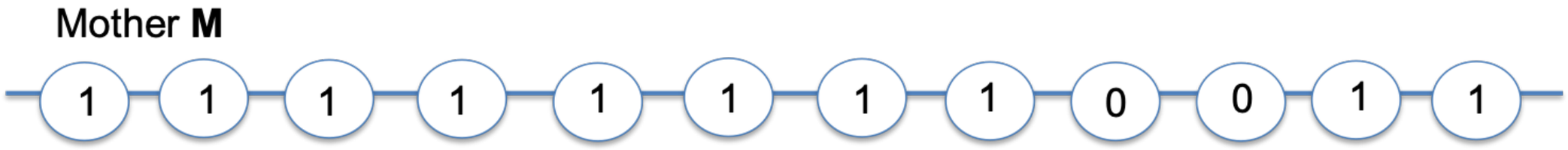
Sibiraj Pillai
Electrical Engineering, IITB

Nithya Ramakrishnan
Postdoctoral fellow, IITB

Known some statistical information about the mother-chromatin, what is the best any known algorithm can do?

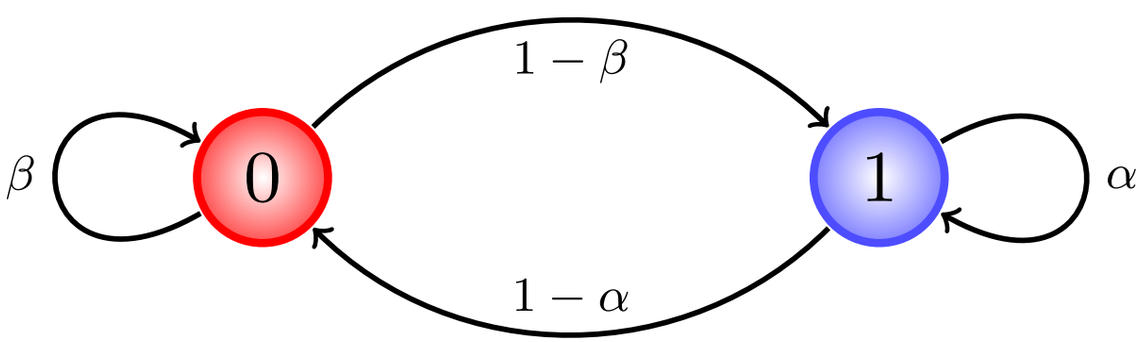


Assume: a mother sequence can be modelled as a Markov chain (of order 1).



Let α be the probability of having a 1 followed by a 1

β be the probability of having a 0 followed by a 0



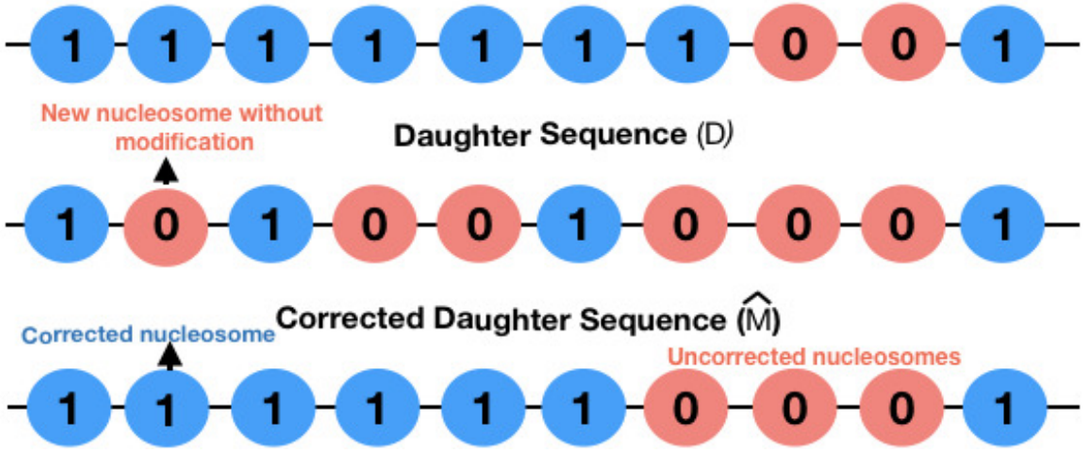
(b)

Some experimentally known sequences of H3k27me3 can be obtained by taking

$$\alpha \approx 0.8; \quad \beta \approx 0.4$$

https://arxiv.org/abs/2005.06539

Given the Markov chain assumption, we use MAP (maximum a posteriori probability) decoding algorithm in communication theory to reconstruct the mother sequence



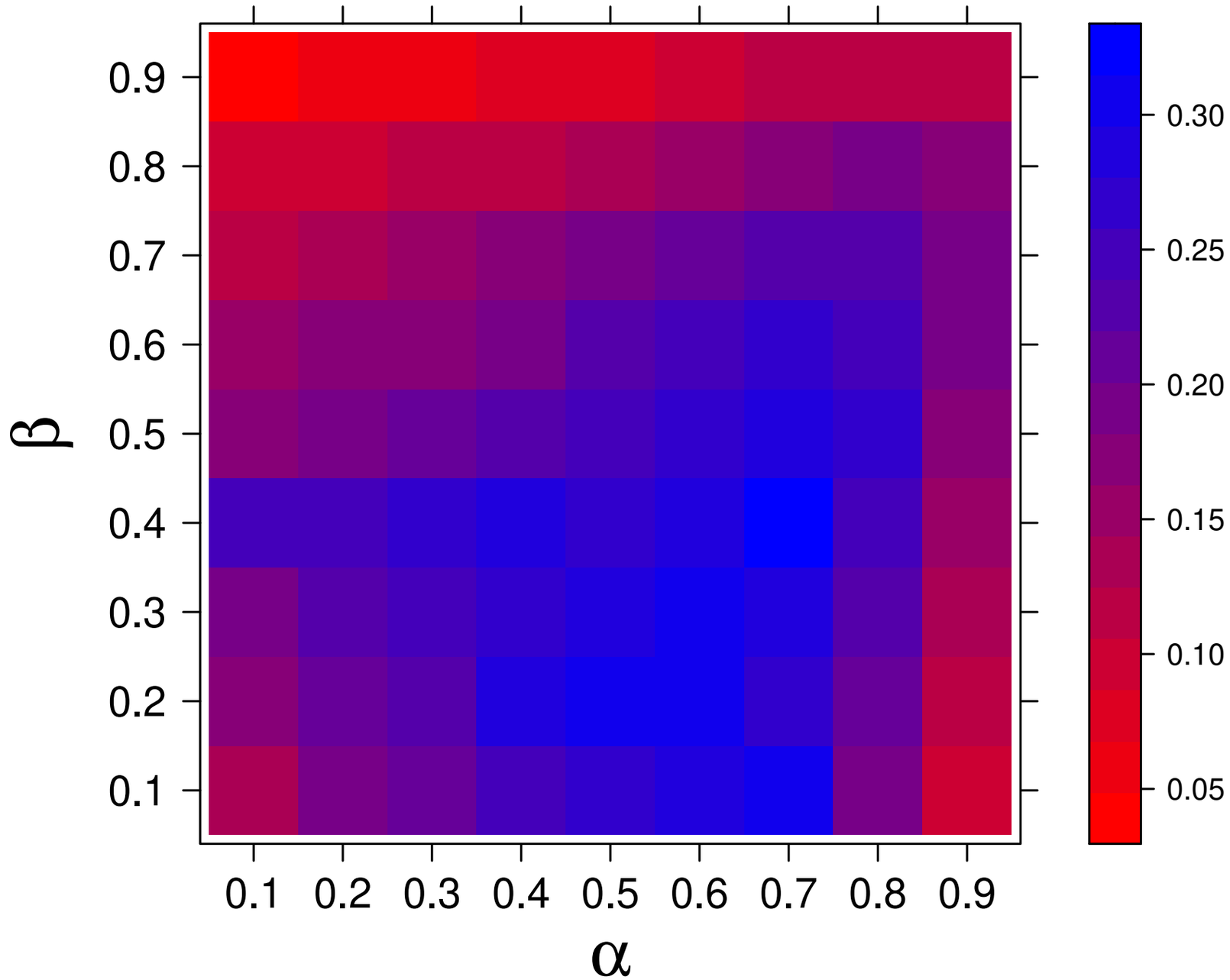
$$(\hat{m}_1, \cdots, \hat{m}_N) = \underset{m_1, \cdots, m_N}{\operatorname{argmax}} \mathbb{P}(m_1^N | d_1^N).$$

We compute the deviation of the re-constructed mother-like sequence from the from the original mother sequence as “error”

$$\Delta(\mathbf{M}, \hat{\mathbf{M}}) = \frac{1}{N} \sum_{i=1}^{N} (m_i - \hat{m}_i)^2.$$

https://arxiv.org/abs/2005.06539

Deviation between original mother sequence and reconstructed mother sequence



https://arxiv.org/abs/2005.06539

MAP-decoding algorithm is probably too complex for simple enzymes!

Can there be a simpler “algorithm”?

https://arxiv.org/abs/2005.06539

We show that in a certain biologically relevant parameter regime (alpha, beta), the MAP algorithm is essentially the same

as

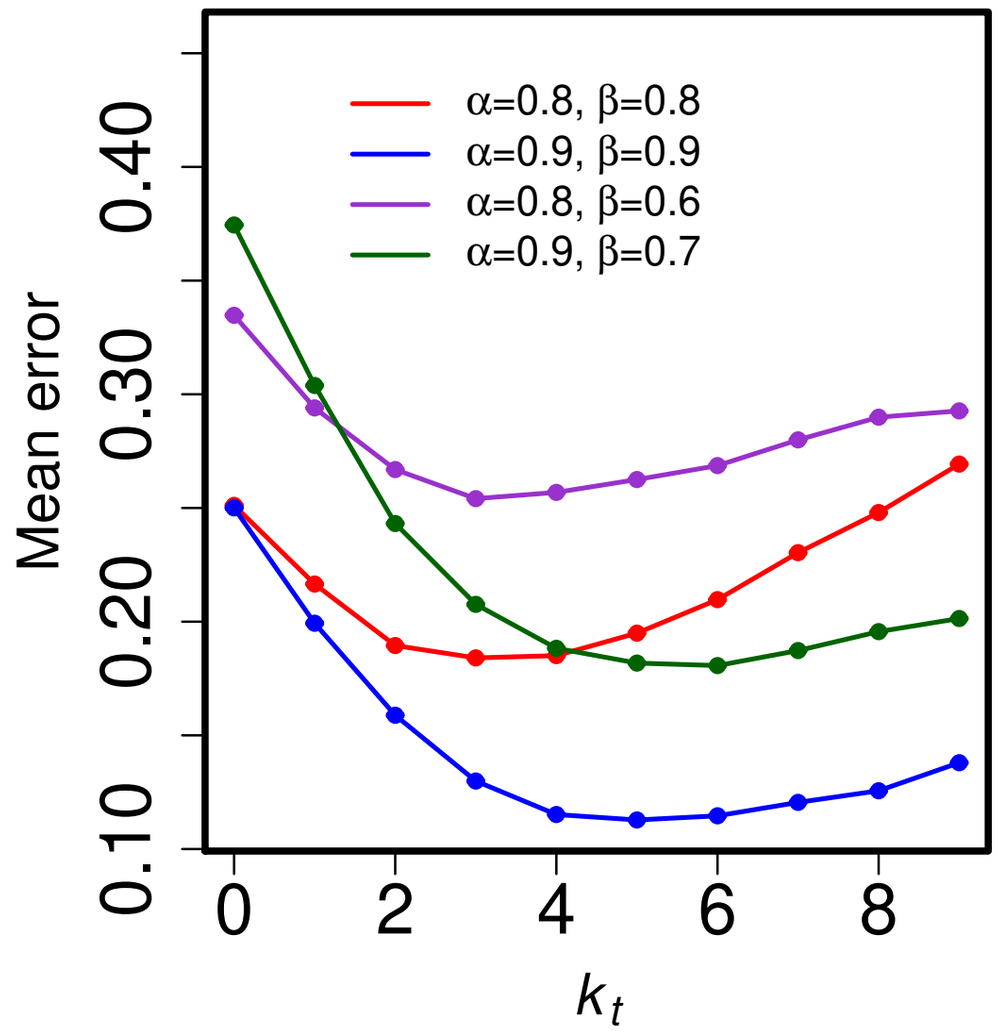
filling “islands” of 0s of size $\leq k$

1111000111100111111000000000111111011111110000000011

Islands with k=3

Fill all "islands" of 0s of size $\leq k$

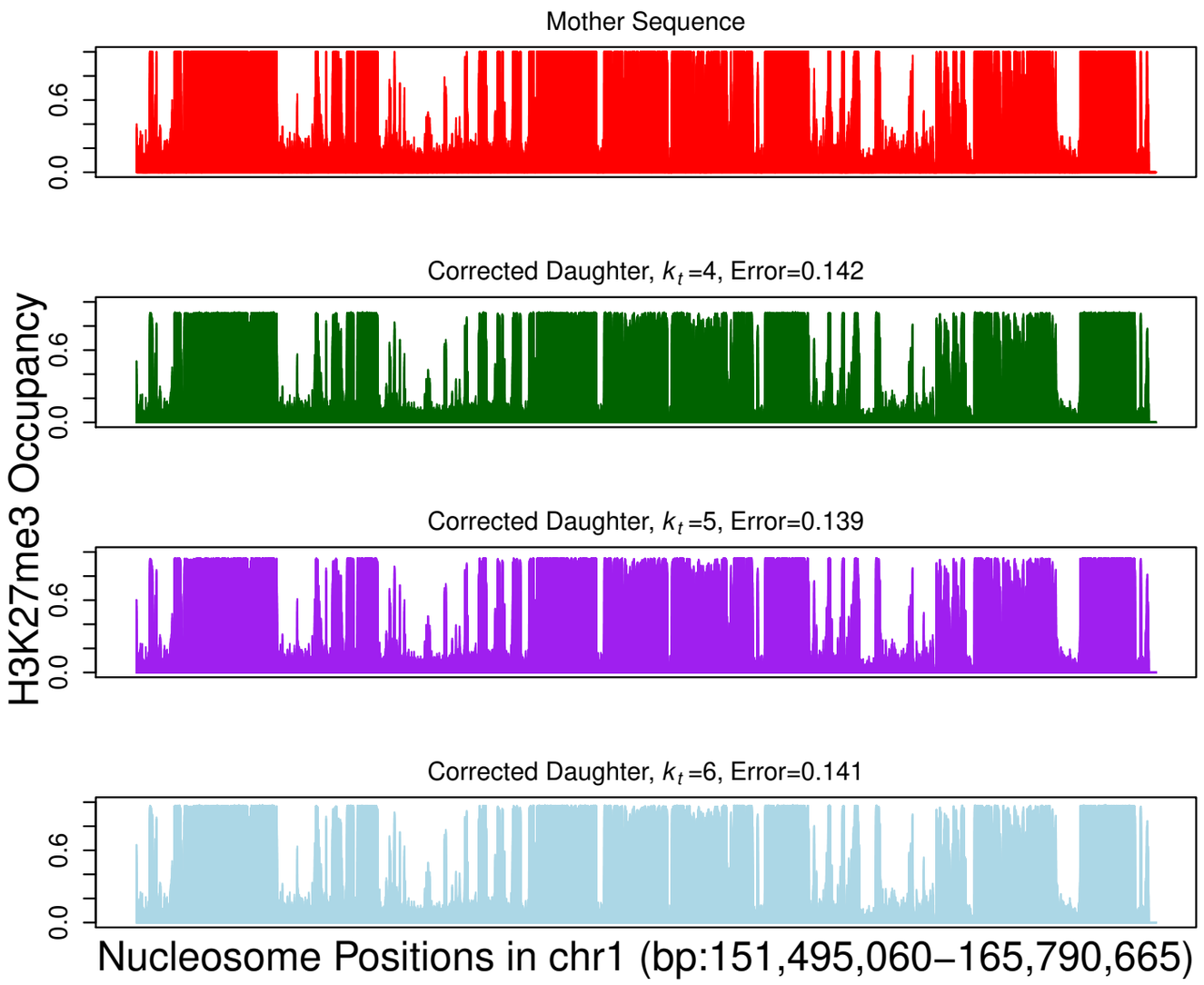
1111000111100111111000000000111111011111110000000011



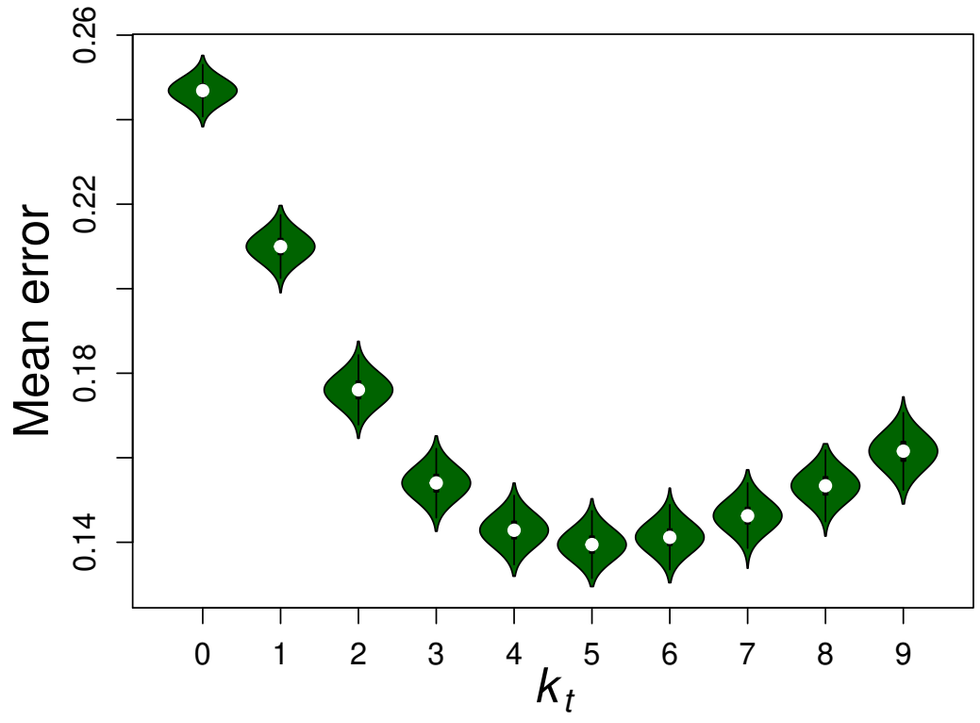
$$k^* = \frac{\log\left(\frac{(1-\alpha)(1-\beta)}{(\alpha^2/2)}\right)}{\log(\alpha/2\beta)} + 1.$$

https://arxiv.org/abs/2005.06539

Our k-filling algorithm can reconstruct experimentally observed patterns



(b)



Experimental data from Groth lab

H3k27me3

https://arxiv.org/abs/2005.06539

Summary-II

- Given partial information about histone modifications after replication, how do cells reconstruct the complete information?

- What can a “machine” do?

- MAP-decoding algorithm = filling islands 0s of size <= k

- Simple enough for an enzyme to execute

- It can reconstruct experimental data reasonably well!

https://arxiv.org/abs/2005.06539

Thank you

Funding: DST/SERB, DBT, IIT Bombay

http://www.bio.iitb.ac.in/~ranjith/

Kiran Kumari
Ravi Jagadeeshan, Monash University
D. Burkhard, Max Planck Institute for Polymer Research

Nithya Ramakrishnan
Sibiraj Pillai, EE, IITB

Gaurav Bajpai
Tripti Bameta
Dibyendu Das
Mandar Inamdar
Ishutesh Jain

Jyotsna Parmar
John F Marko
Hungyo Kharerin