

Recurrence networks and contact maps: an unified approach to dynamical systems and genomics

Annick LESNE

*CNRS (French National Center for Scientific Research)
LPTMC (Paris) & IGMM (Montpellier)*

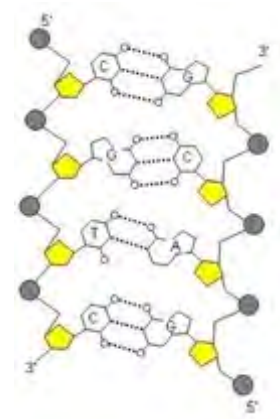
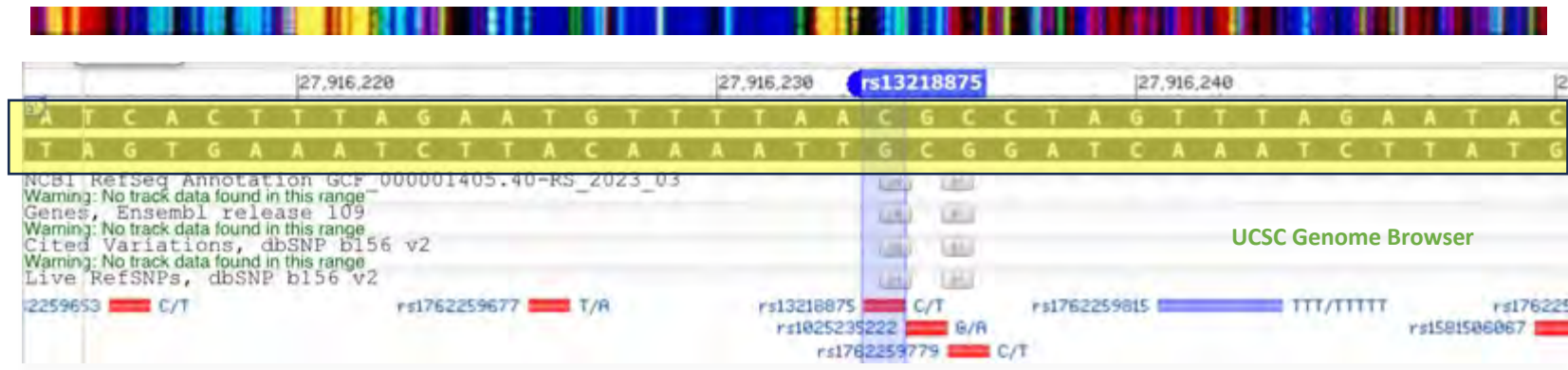
Work done with **Julien MOZZICONACCI** (MNHN, Paris) and **Sofiane RAMDANI** (University of Montpellier, LIRMM)



Human genome : 46 DNA molecules harboring our genetic heritage

Nature 2021

1D view



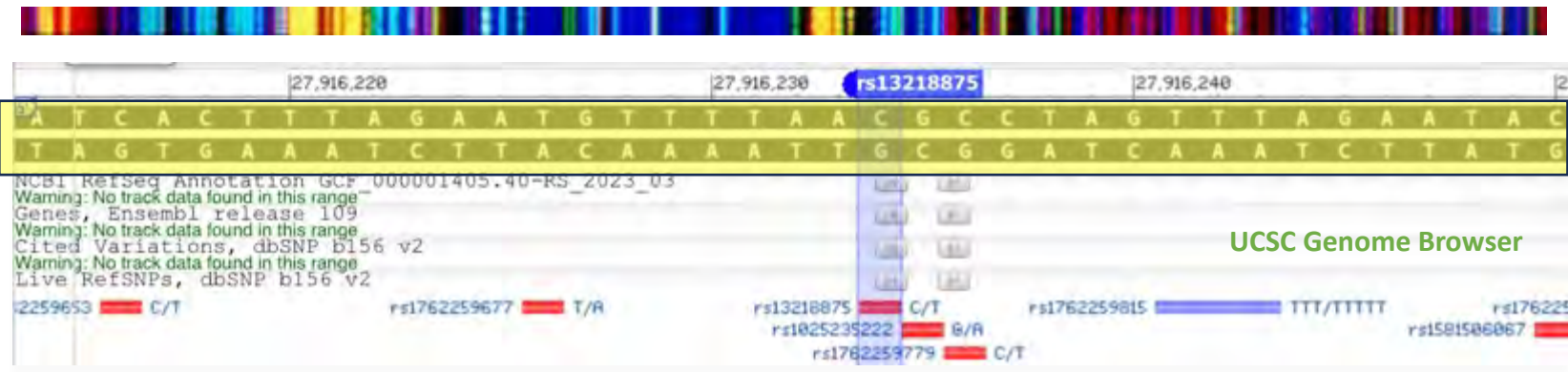
- A - T
- A - T
- G - C
- C - G
- C - G
- T - A
- T - A
- A - T
- C - G

views of the DNA molecule at different scales

Human genome : 46 DNA molecules harboring our genetic heritage

Nature 2021

1D view

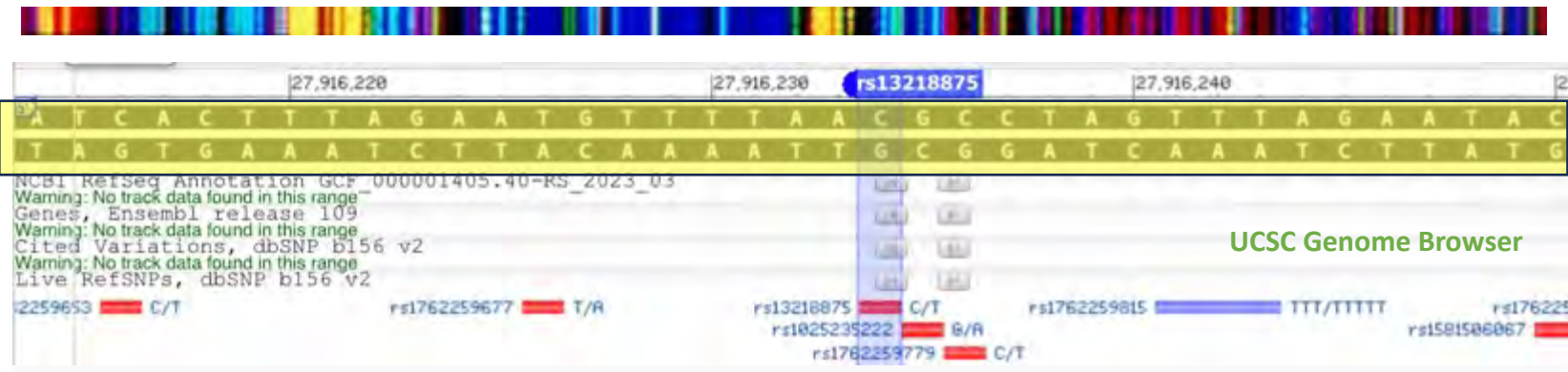


sequence analysis: position along the genome ↔ time

Human genome : 46 DNA molecules harboring our genetic heritage

Nature 2021

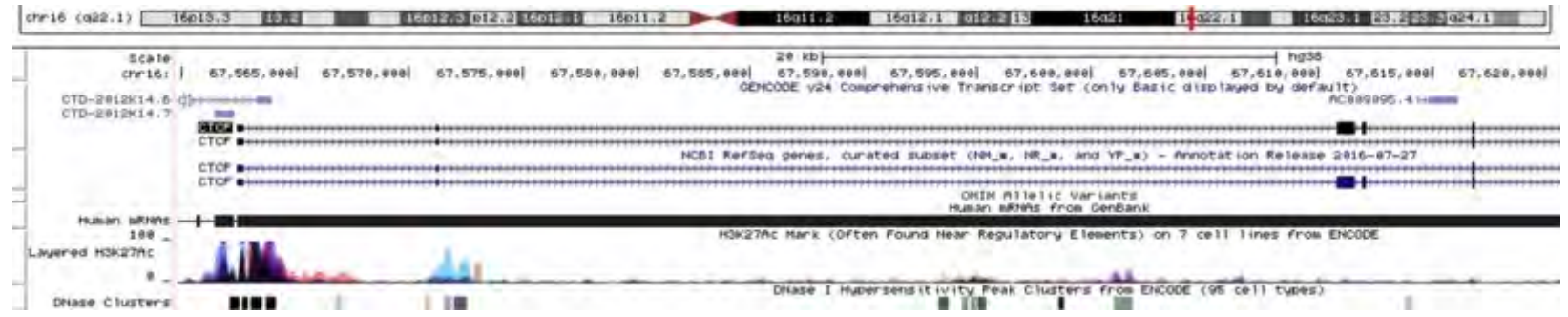
1D view



UCSC Genome Browser

sequence analysis: position along the genome ↔ time

annotated sequence



UCSC Genome Browser

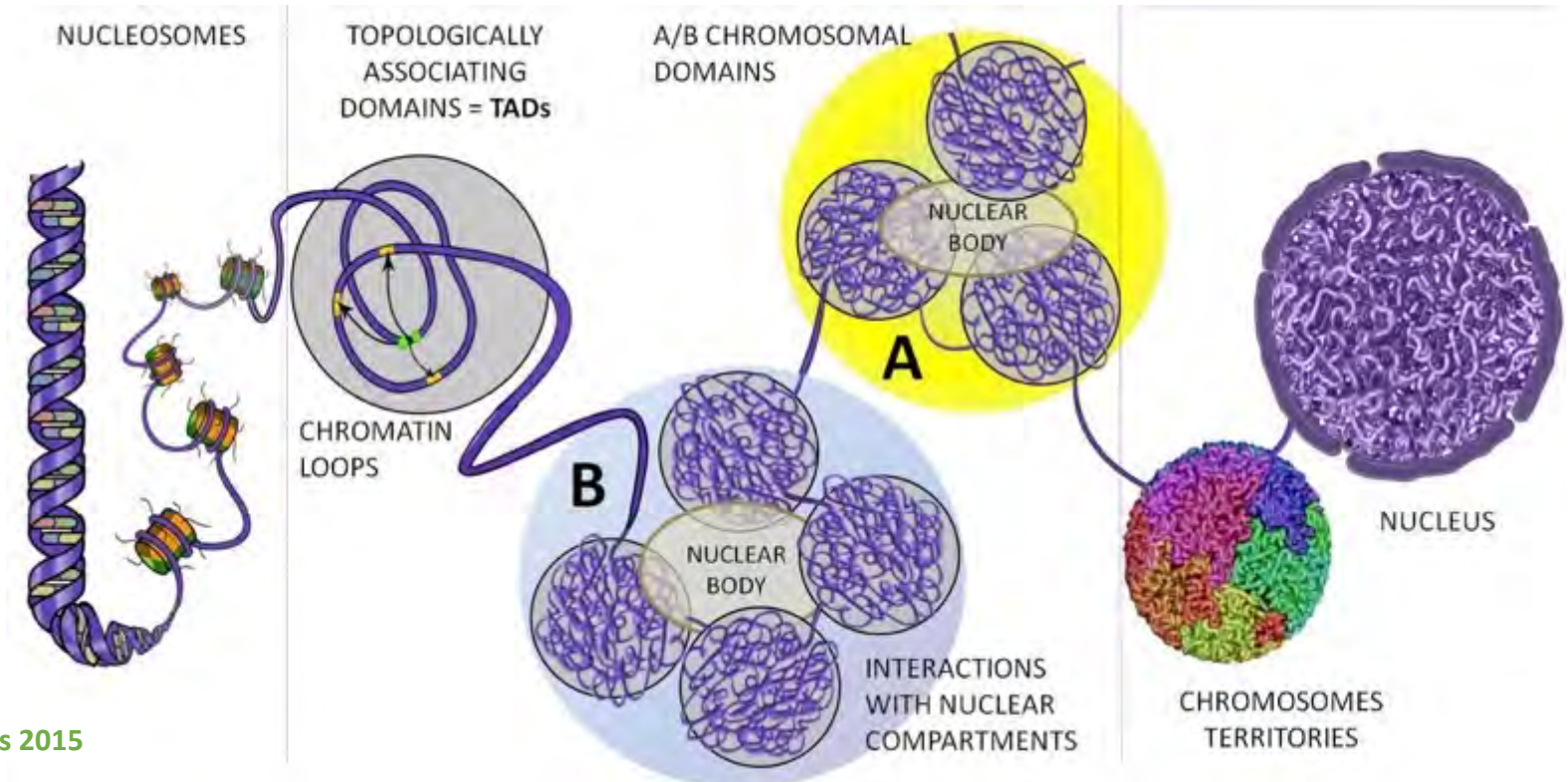
Genome: 1D vs 3D view

Nature 2021

1D view



3D view

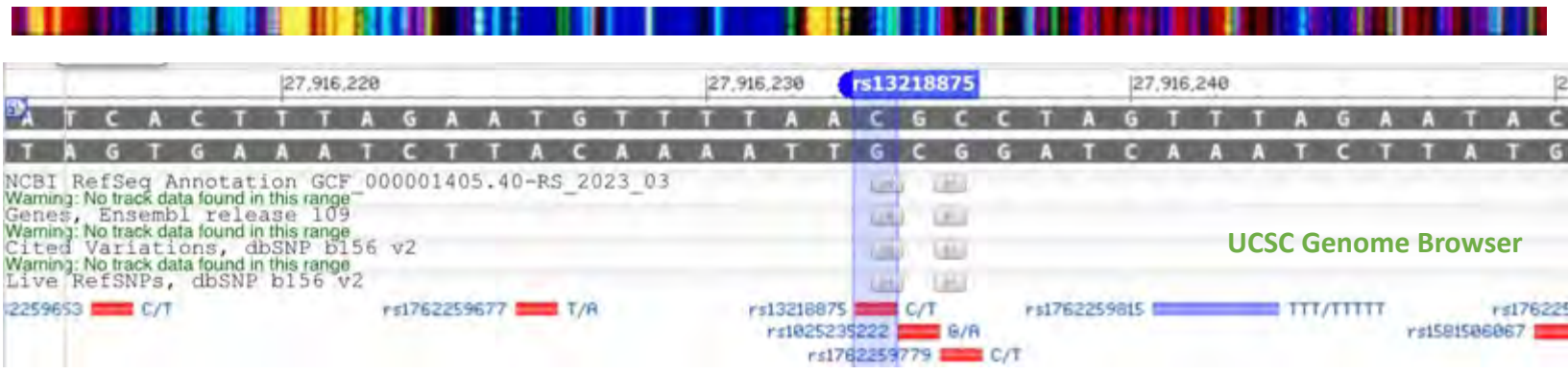


Ea et al. Genes 2015

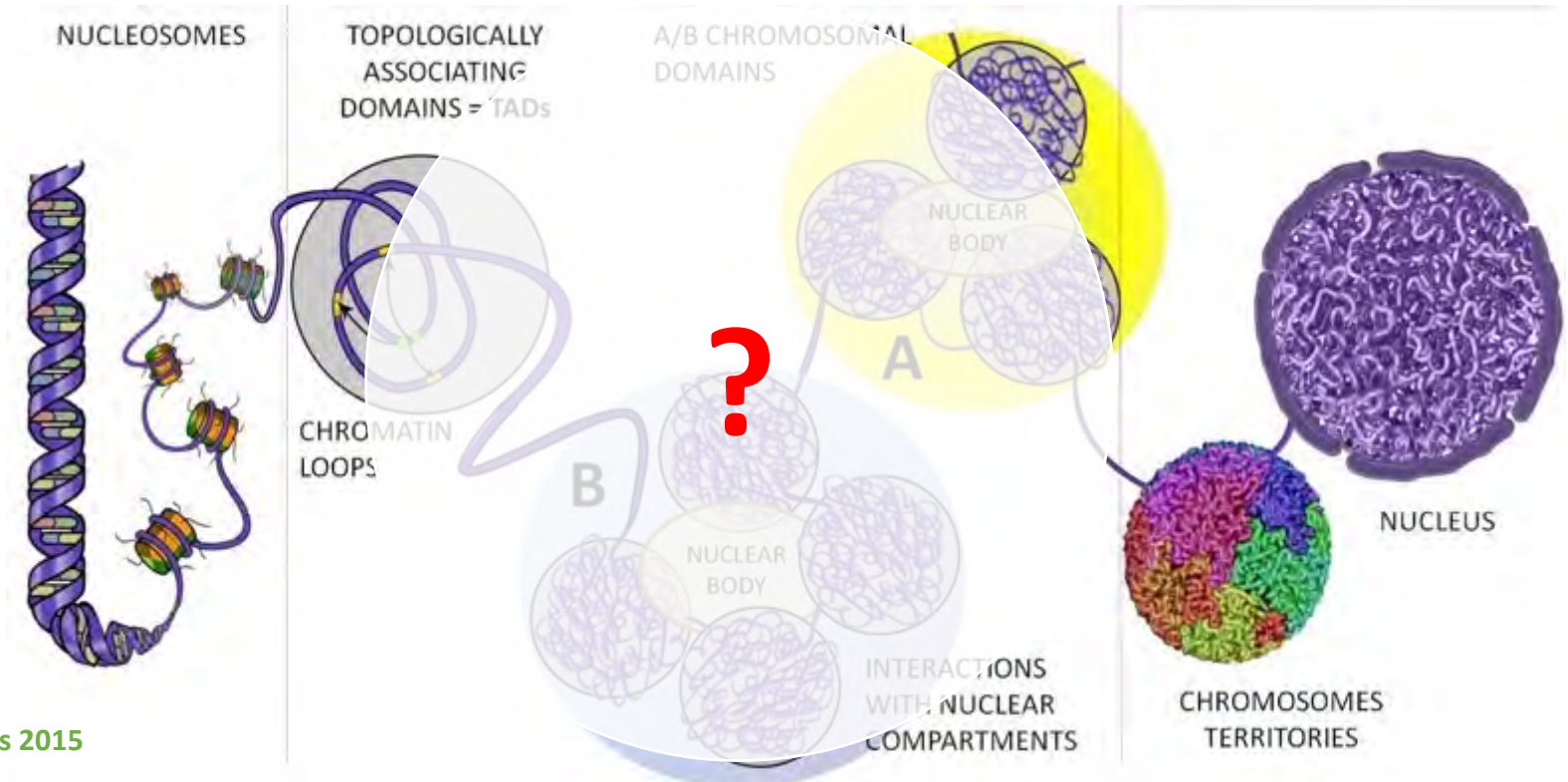
Genome: 1D vs 3D view

Nature 2021

1D view

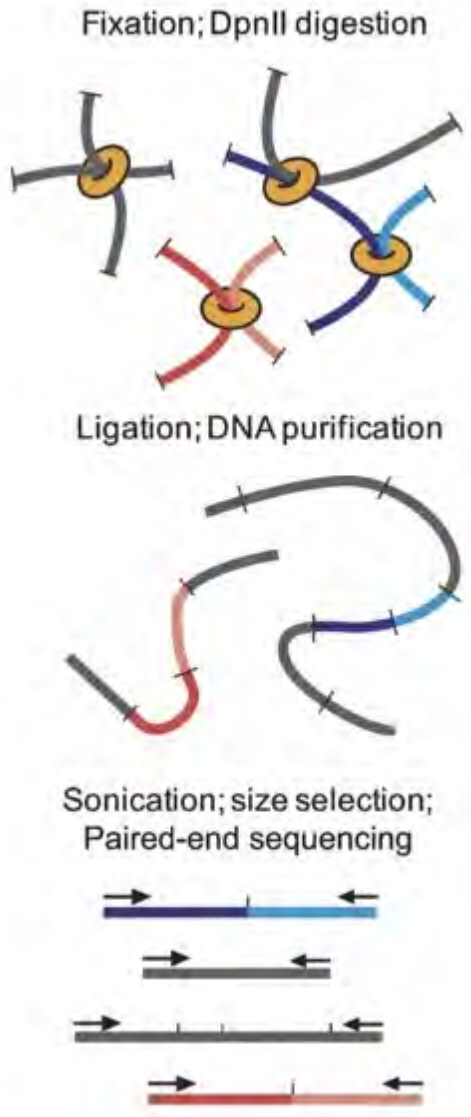


3D view

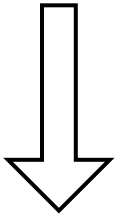


Ea et al. Genes 2015

Chromosome Conformation Capture (3C): the experimental technique



3D genome conformation in vivo
(living cells)

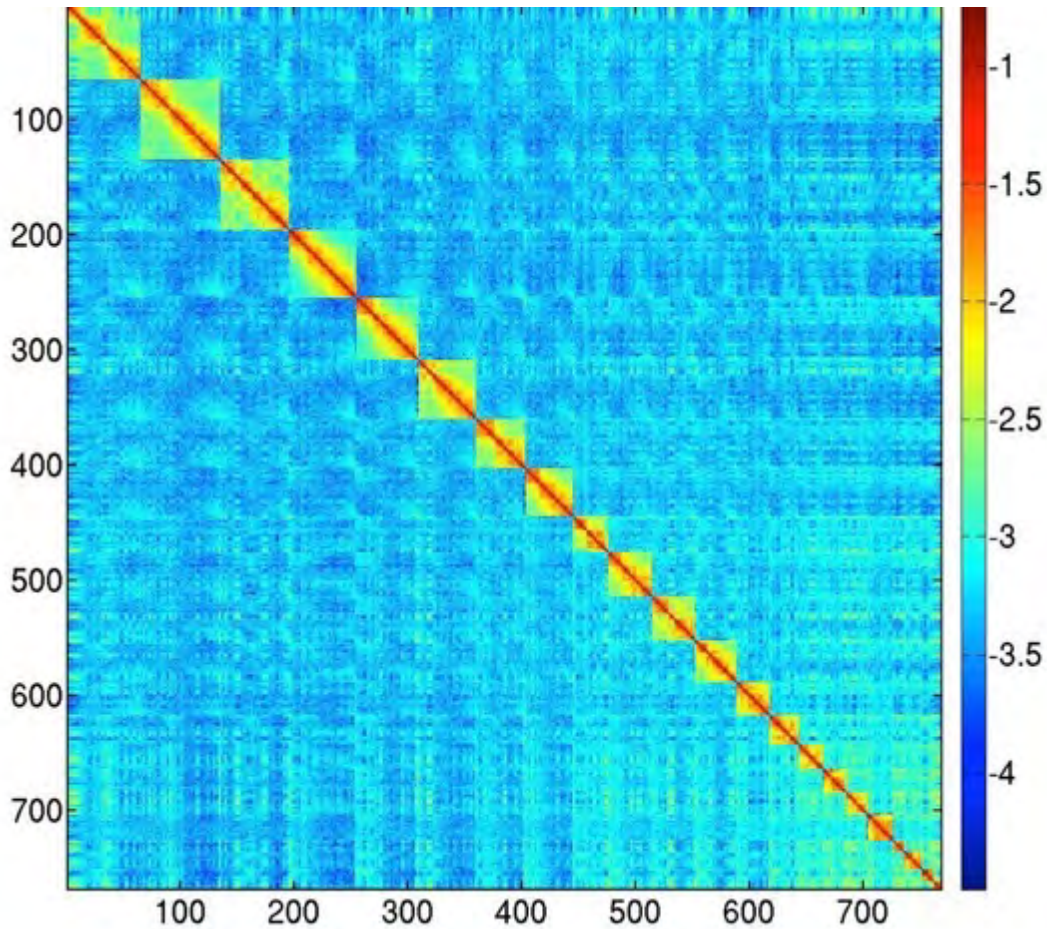
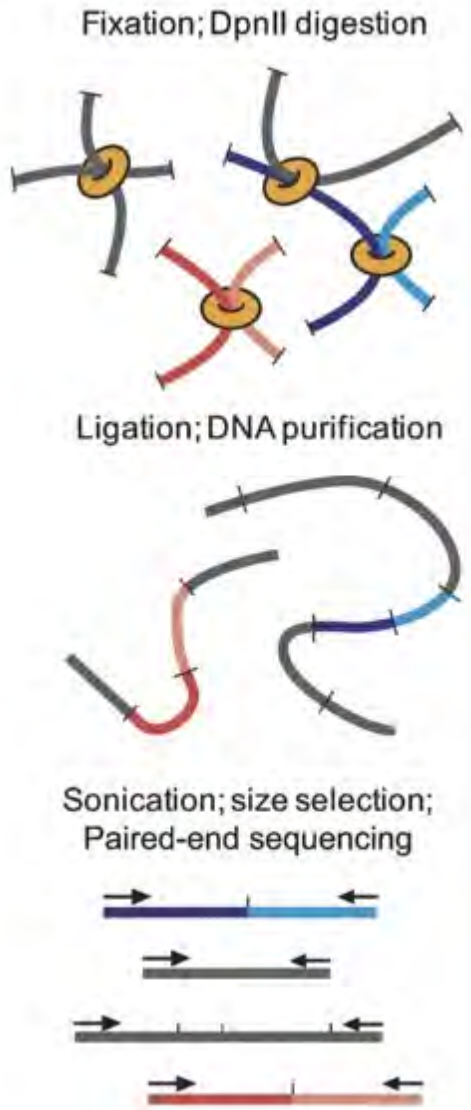


'contacts'
(pairs of genomic loci close in the 3D space)

+ data processing
(filtering out unreliable/biased reads)

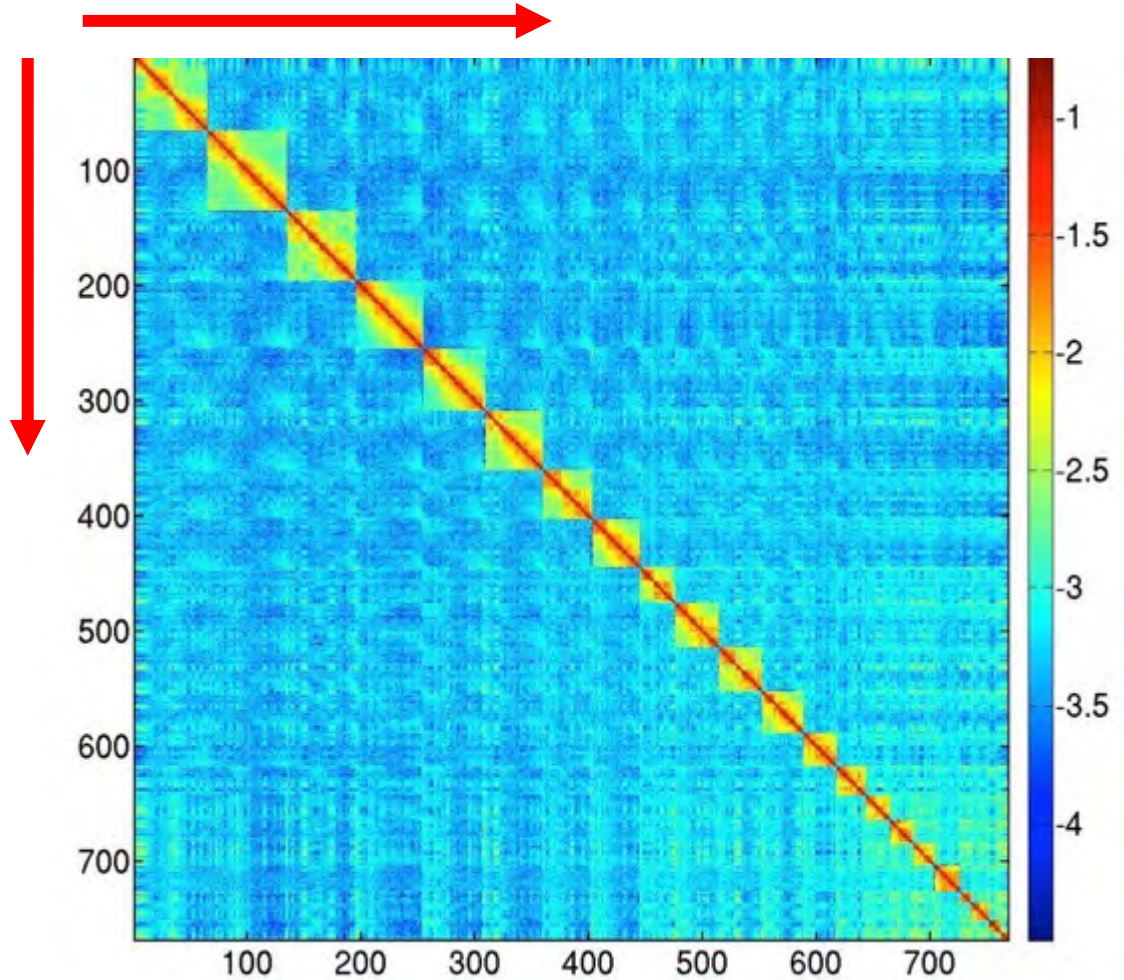
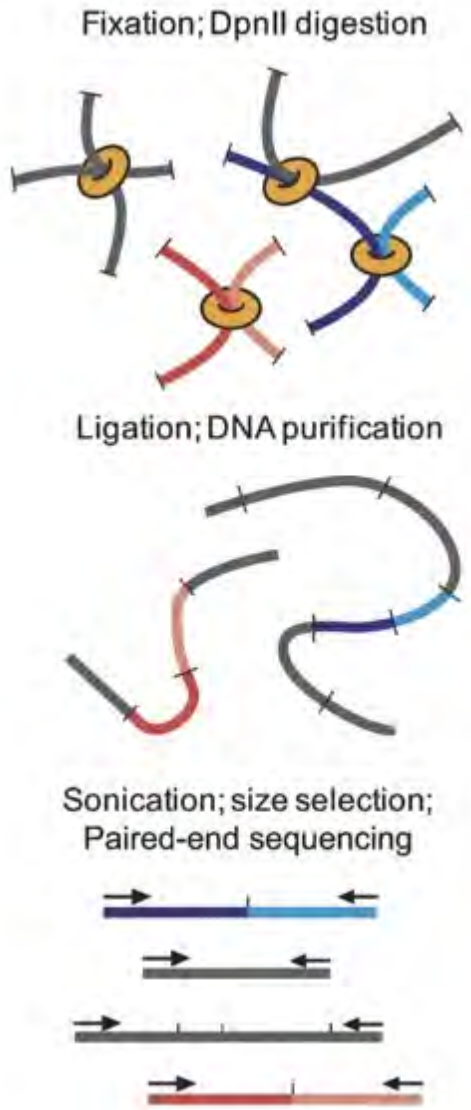
Chromosome conformation capture: Hi-C maps

Hi-C = High-throughput 3C



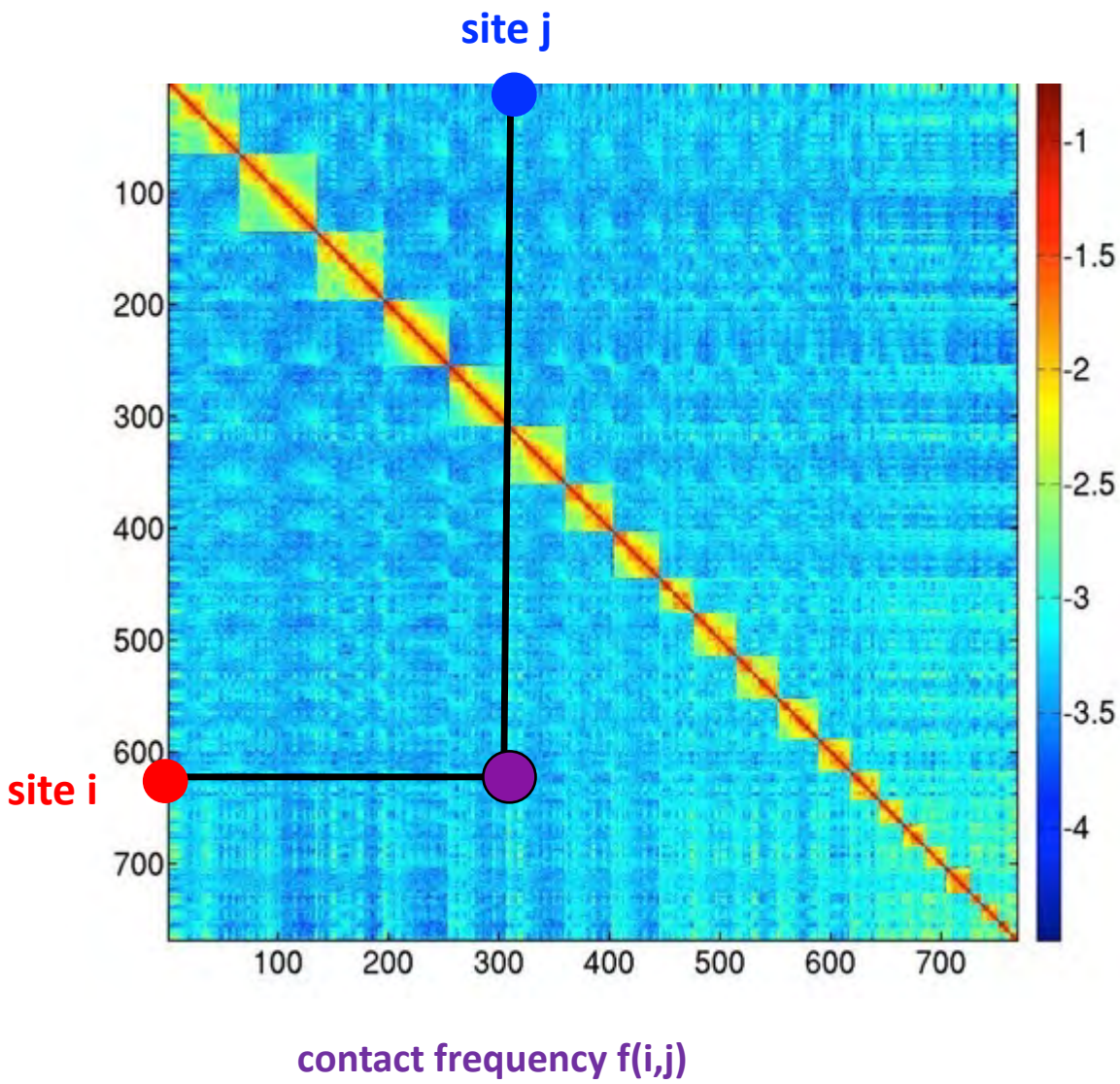
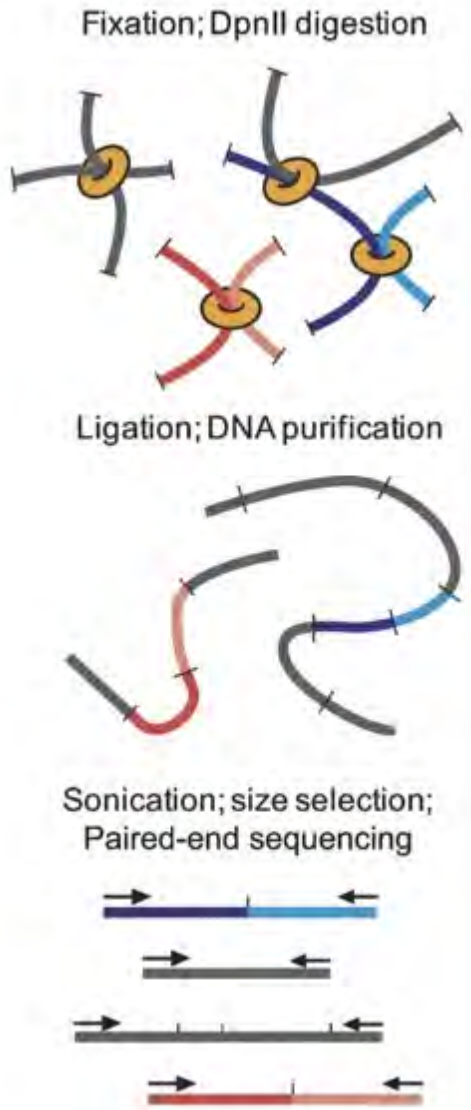
Lesne et al. 2014
Human cells (hESC)

Chromosome conformation capture: Hi-C maps

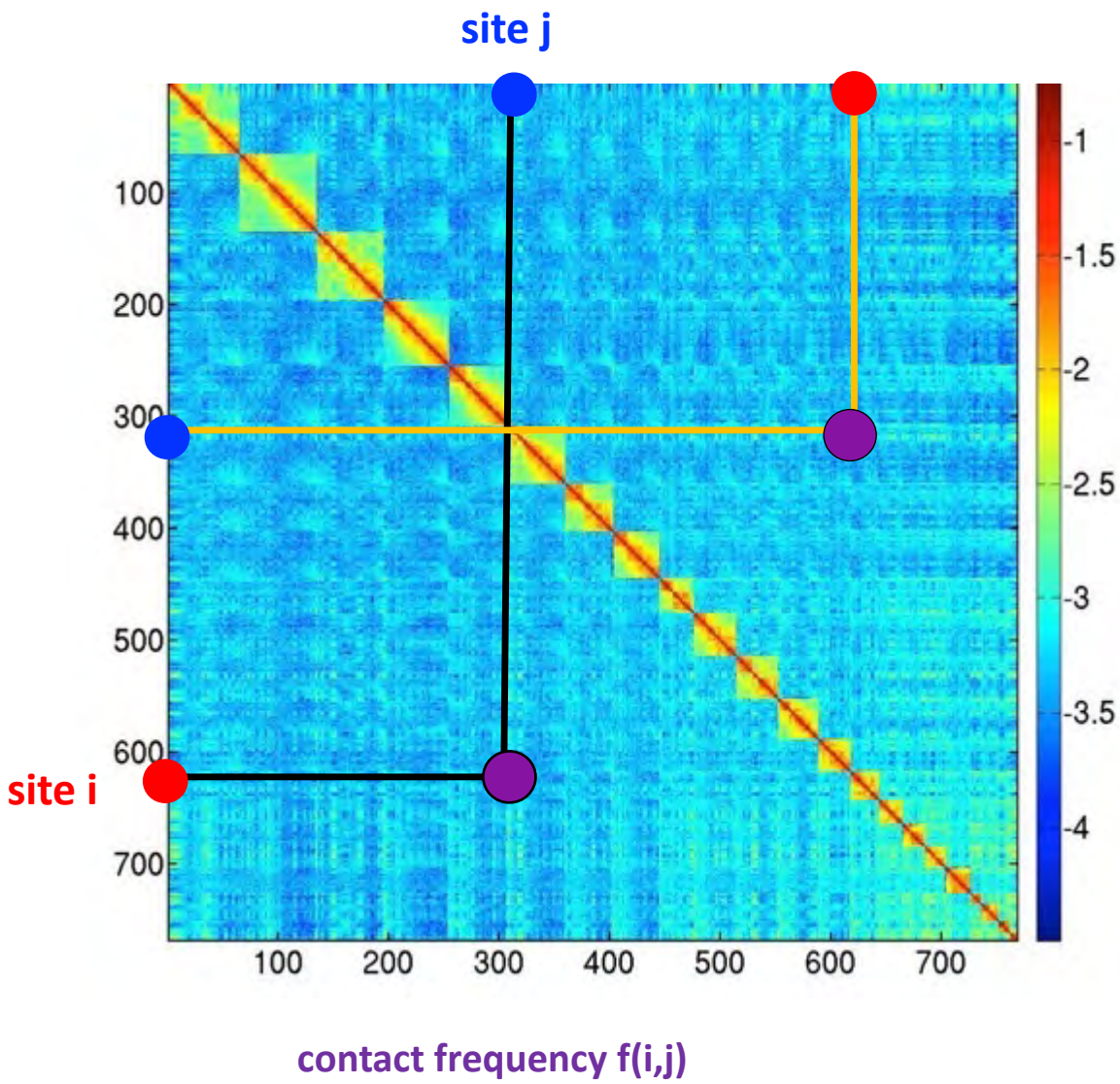
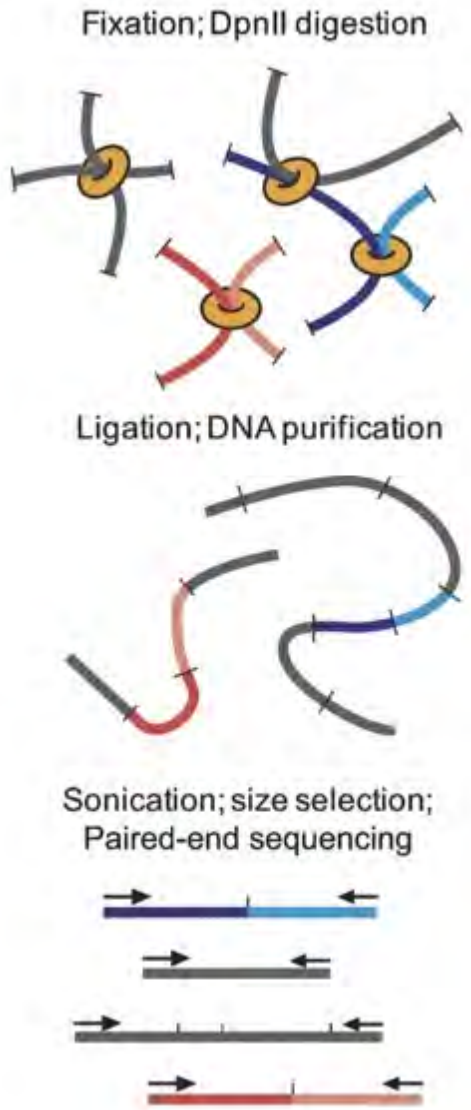


Lesne et al. 2014
Human cells (hESC)

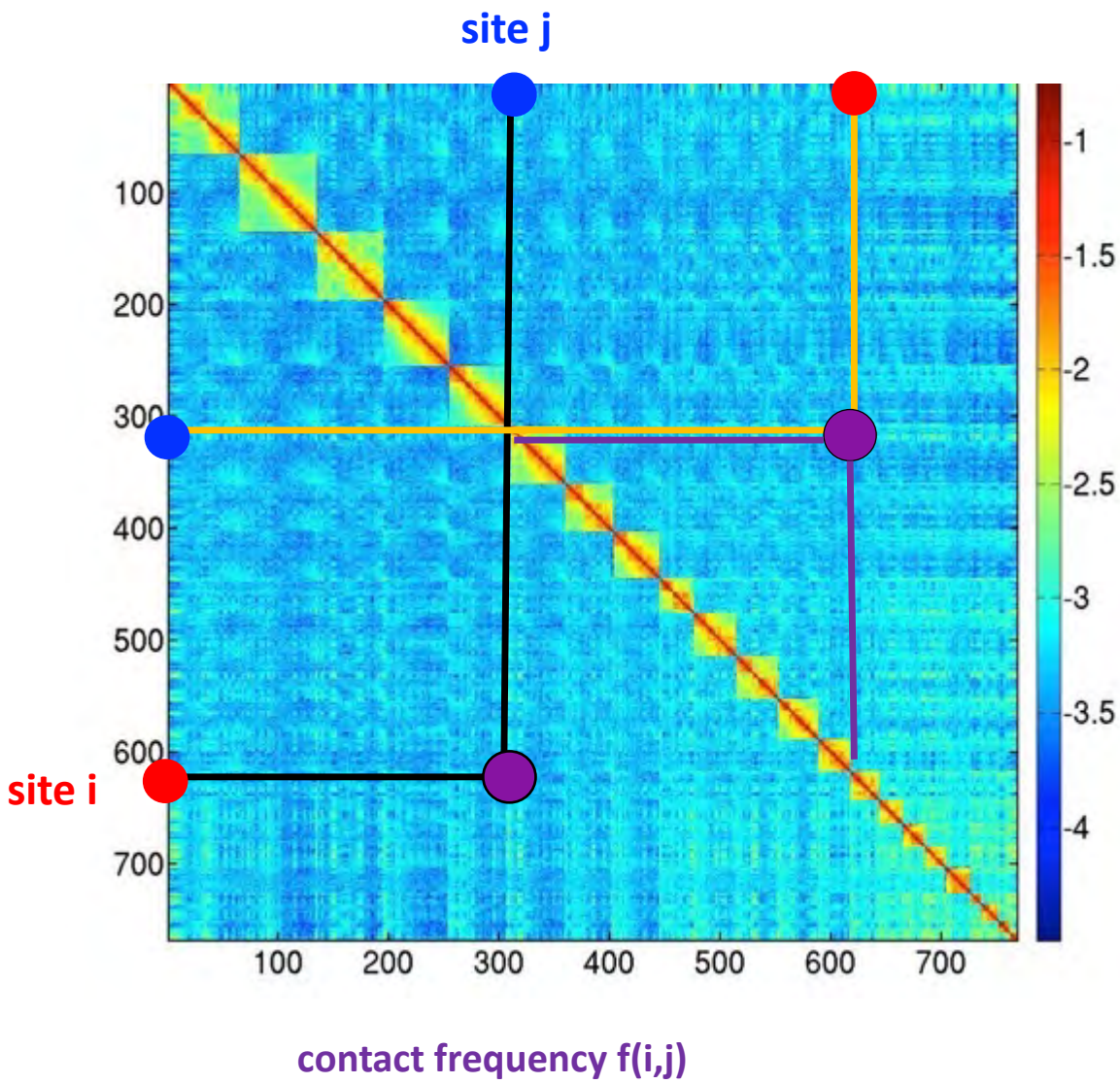
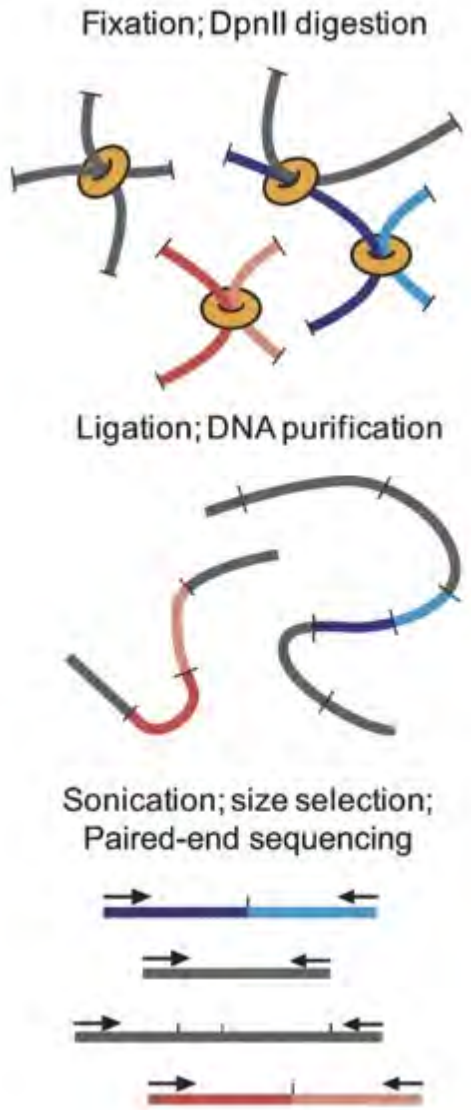
How to read a Hi-C map



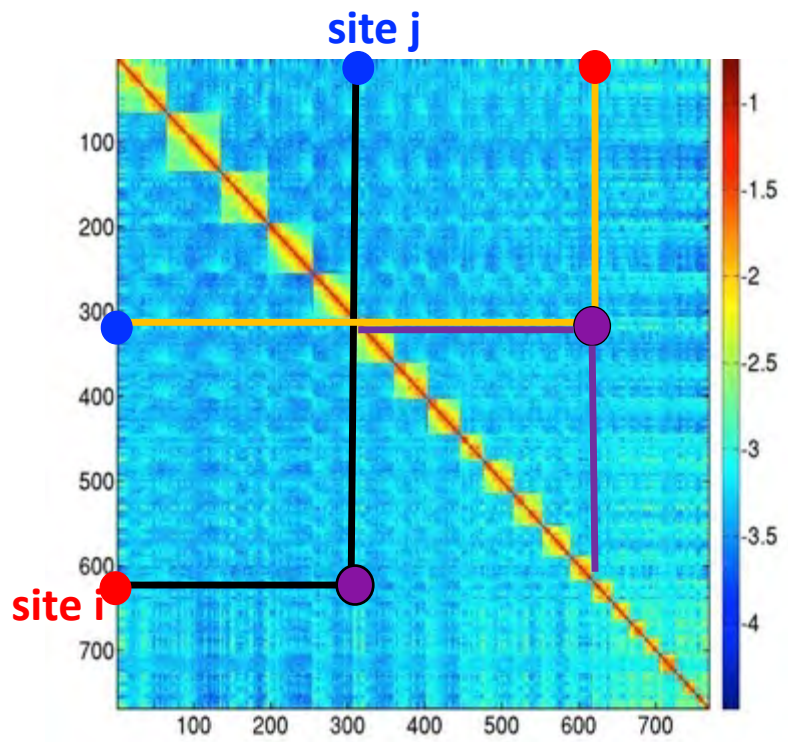
How to read a Hi-C map



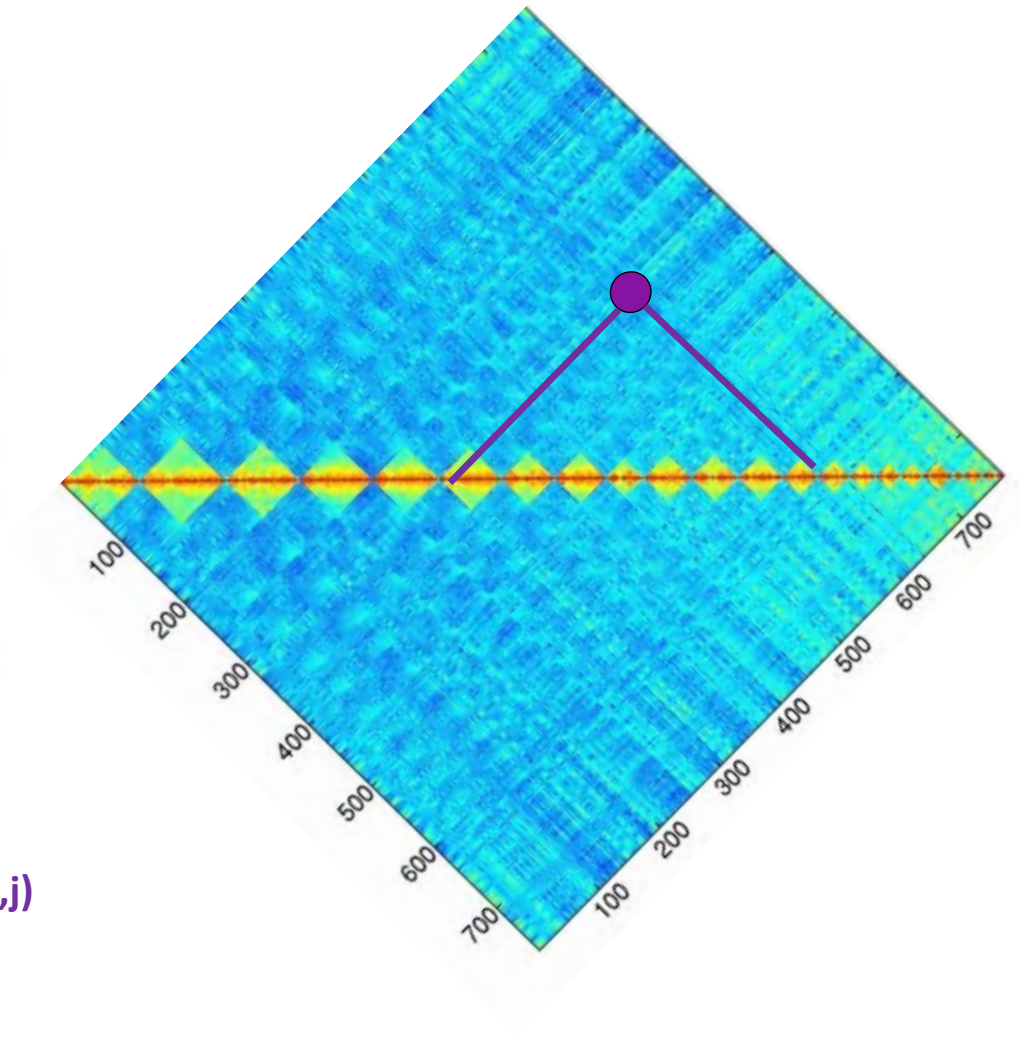
How to read a Hi-C map



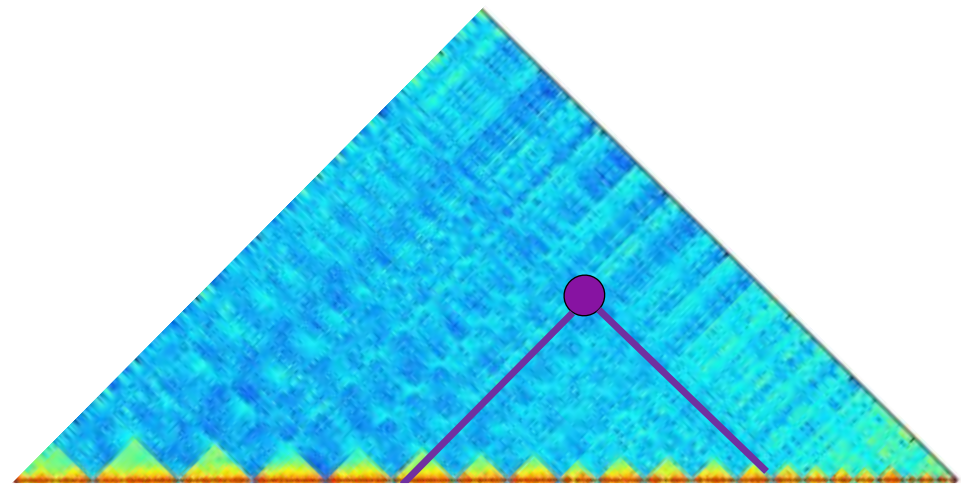
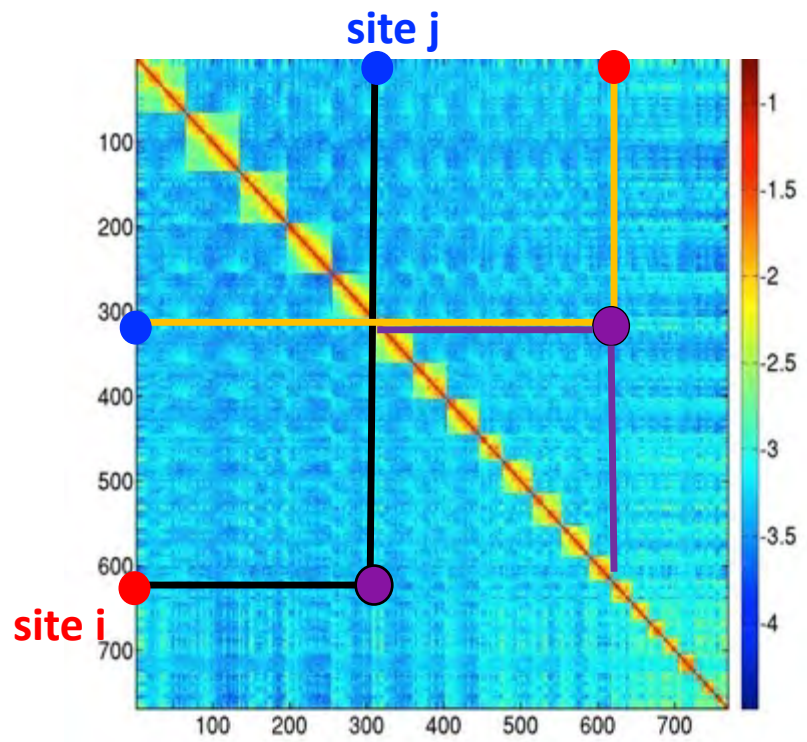
How to read a Hi-C map



contact frequency $f(i,j)$



How to read a Hi-C map

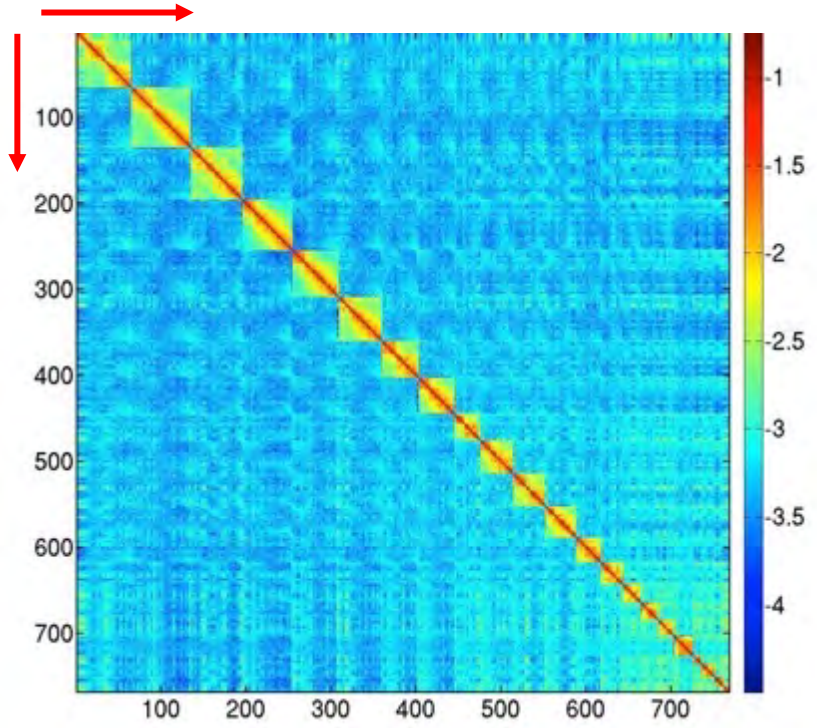


contact frequency $f(i,j)$

Contact maps

$A_{ij} = 1$ means that i and j are in contact (or close in space)

Chromosome conformation



Lesne et al. 2014

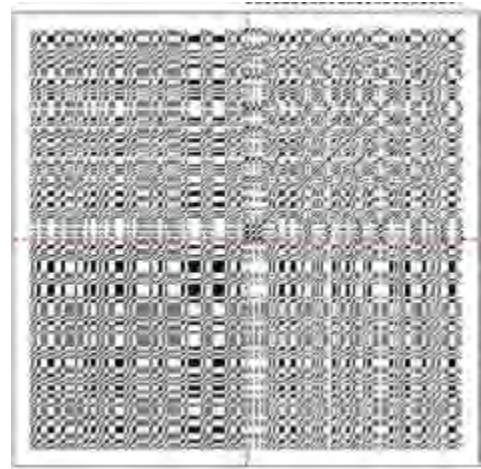
Red = high contact frequency

real space

axes: position along the genome

f_{ij} = contact frequency (real number)

Recurrence plots



Eckmann, Kamphorst & Ruelle EPL 1987

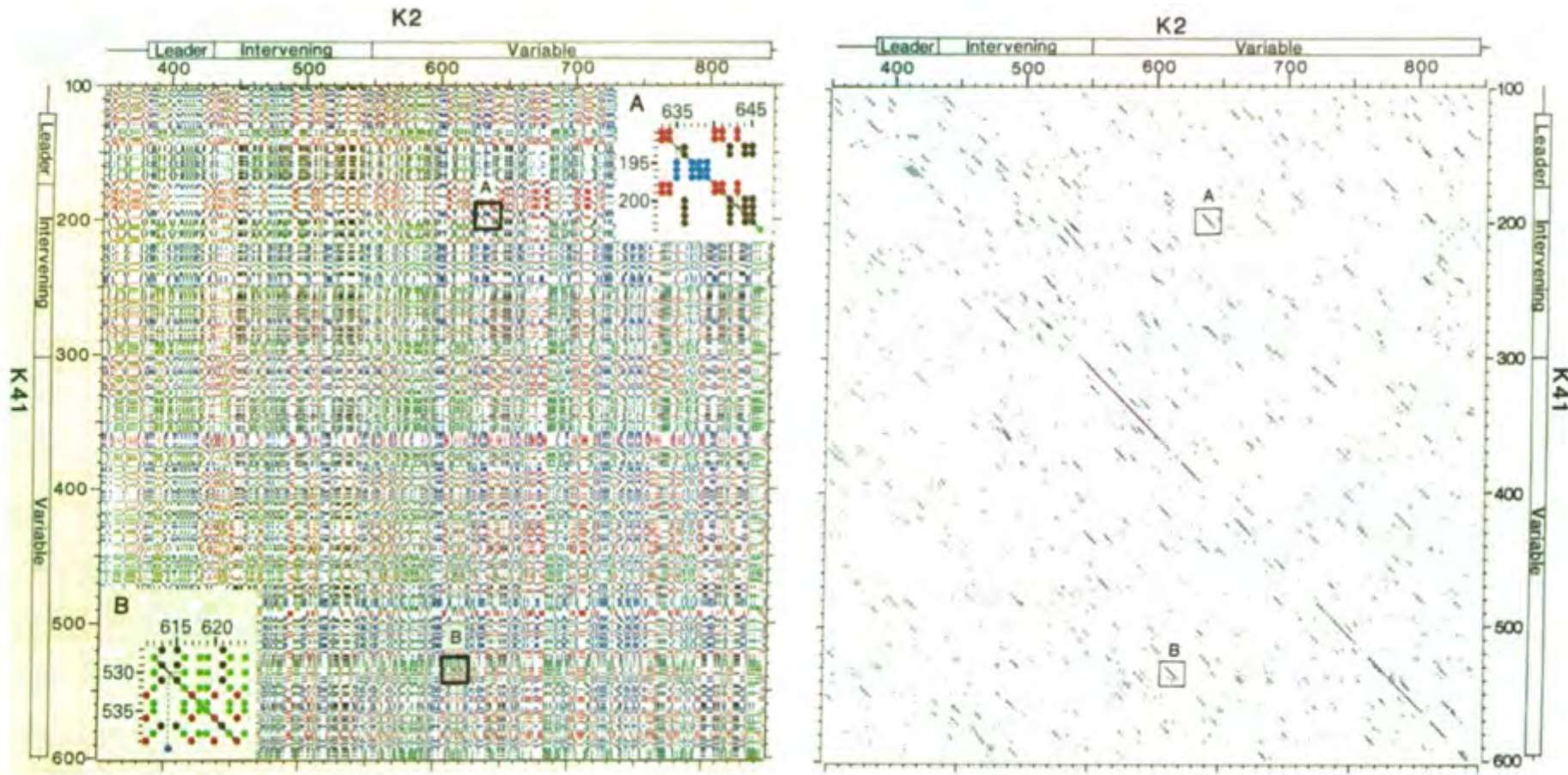
Marwan et al. Physics Reports 2007

phase space

axes: time

Caveat: maps of spatial contacts, not of sequence similarity

Maizel & Lenk, PNAS 1981
comparing two genomic sequences

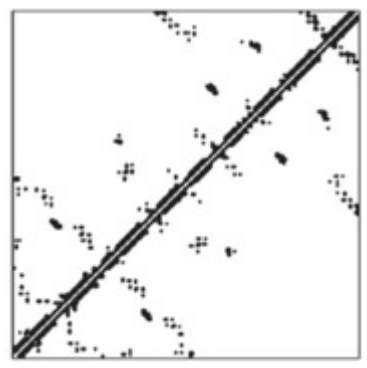


A-A: red T-T: black C-C: green G-G: blue different N-N': empty

Contact maps

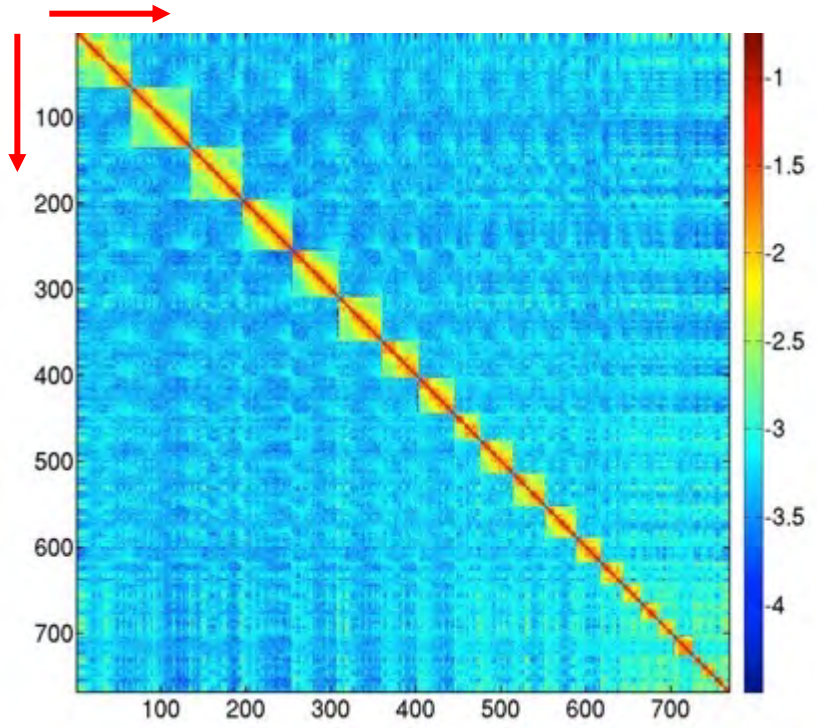
$A_{ij} = 1$ means that i and j are in contact (or close in space)

Protein structure



Colosimo & Giuliani
in *RQA*, Springer 2015
real space
axes: residues

Chromosome conformation

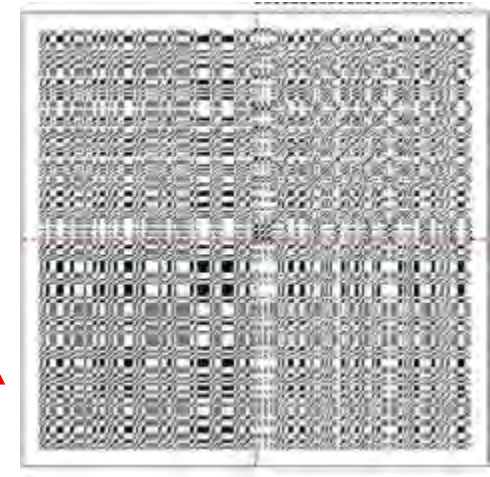


Lesne et al. 2014

Red = high contact frequency

real space
axes: position along the genome
 f_{ij} = contact frequency

Recurrence plots



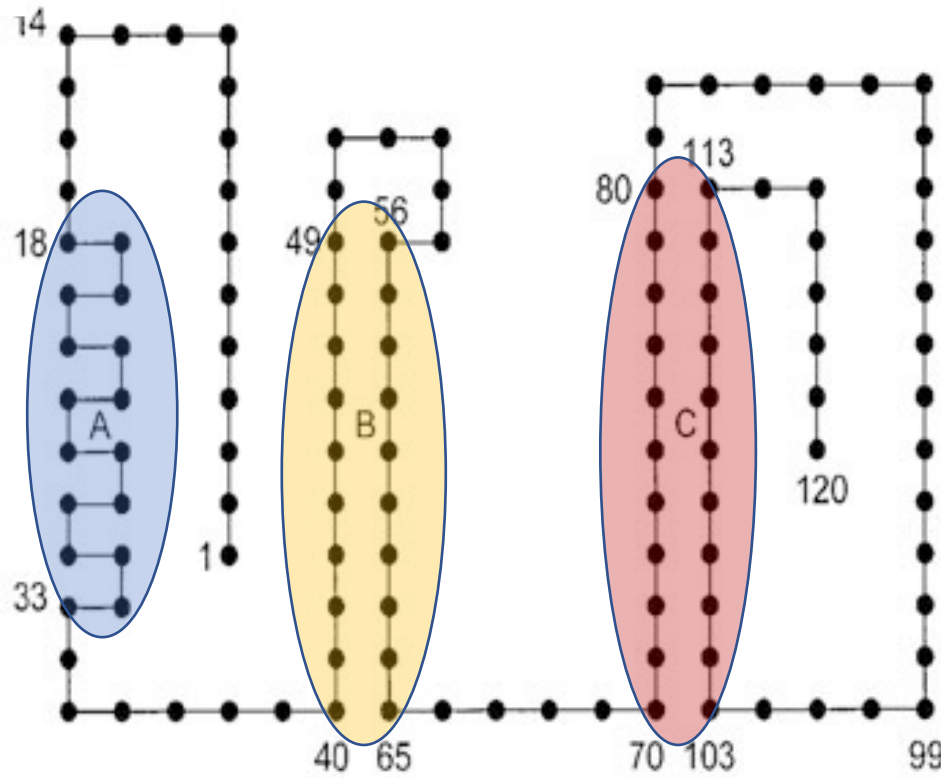
Eckmann, Kamphorst
& Ruelle EPL 1987

Marwan et al.
Physics Reports 2007

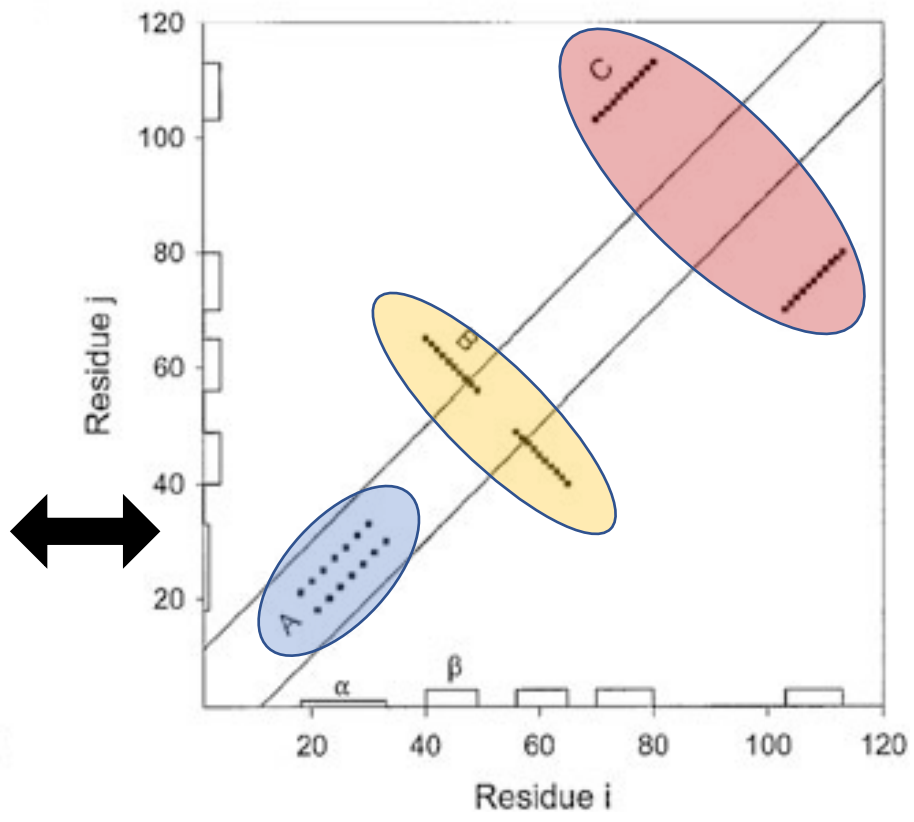
phase space
axes: time

Contact maps: direct analysis for proteins

RP and protein motifs (2D model)



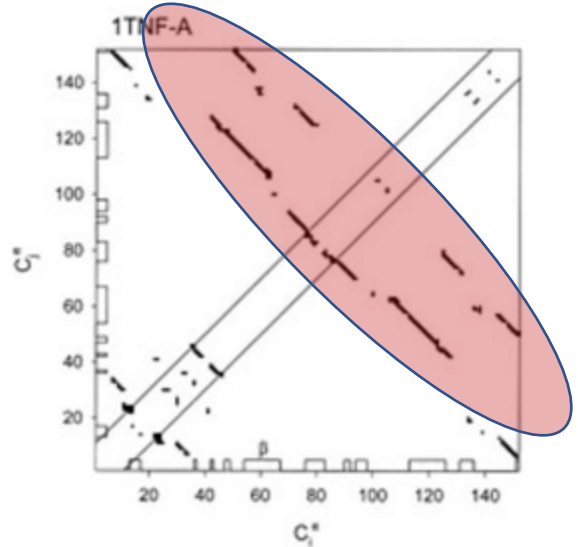
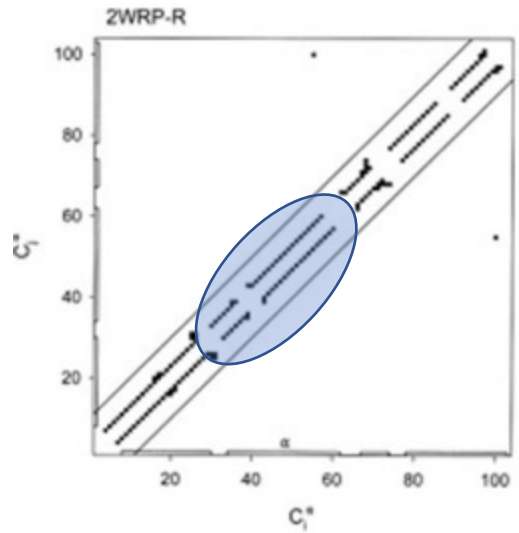
- A: alpha helix
- B: antiparallel beta sheet
- C: parallel beta sheet



adapted from Colosimo & Giuliani in RQA, Springer 2015

Contact maps: direct analysis for proteins

RP and protein motifs (real proteins)



adapted from
Colosimo & Giuliani
In RQA, Springer 2015

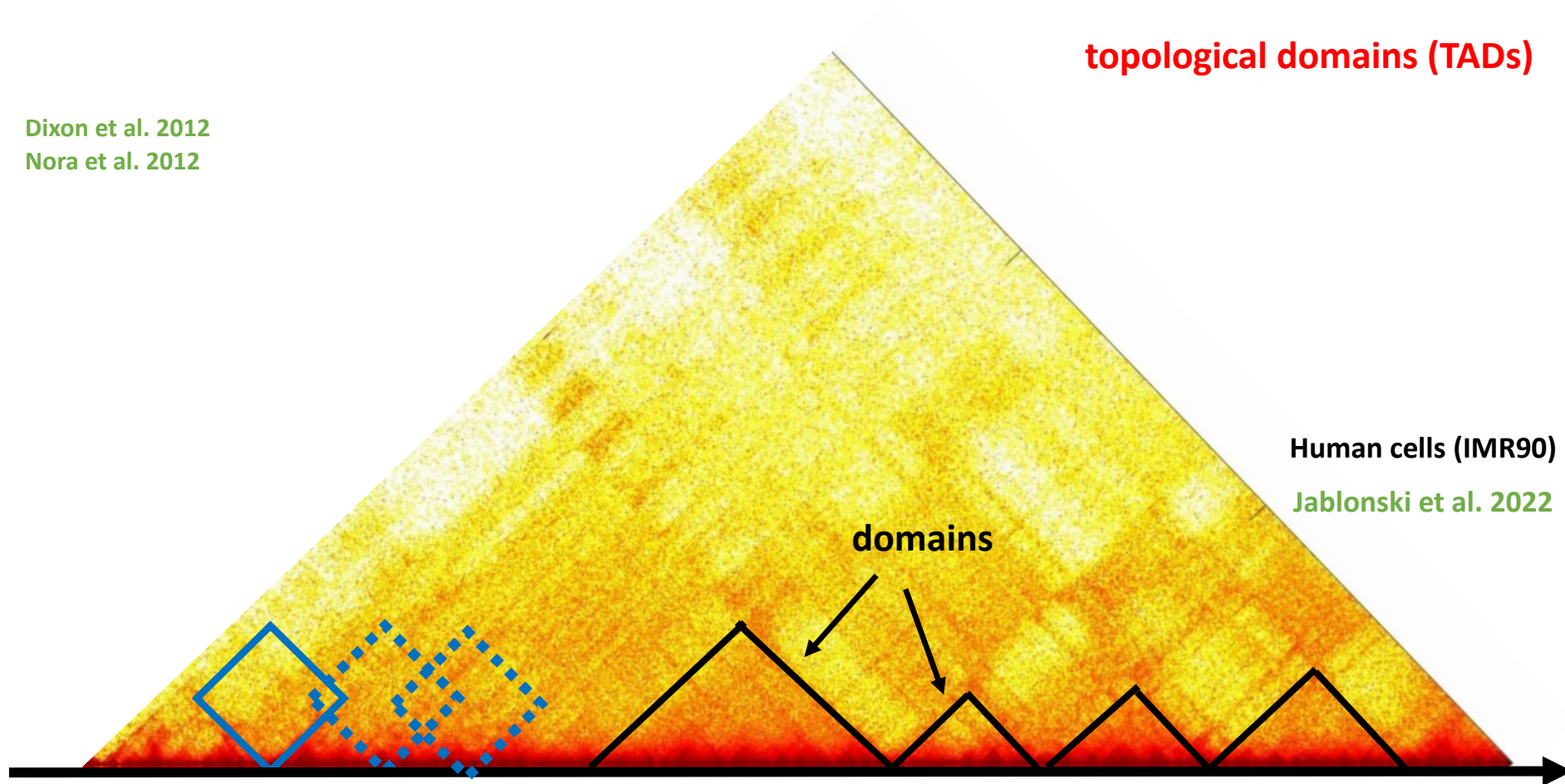
2WRP-R
mainly **alpha helix**

1TNF-A
mainly **beta sheets**

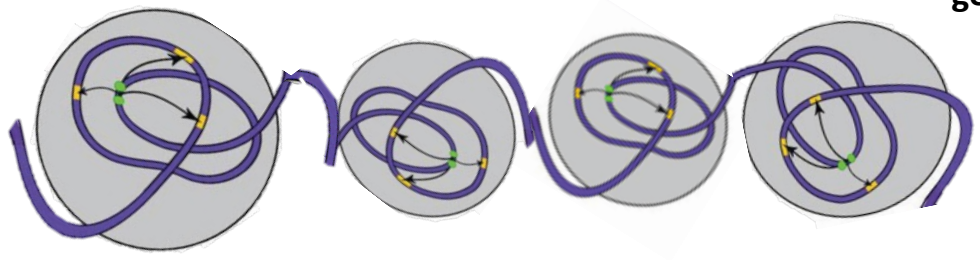
Contact maps: direct analysis for chromosomes

Dixon et al. 2012
Nora et al. 2012

topological domains (TADs)



various measures and algorithms to delineate these domains

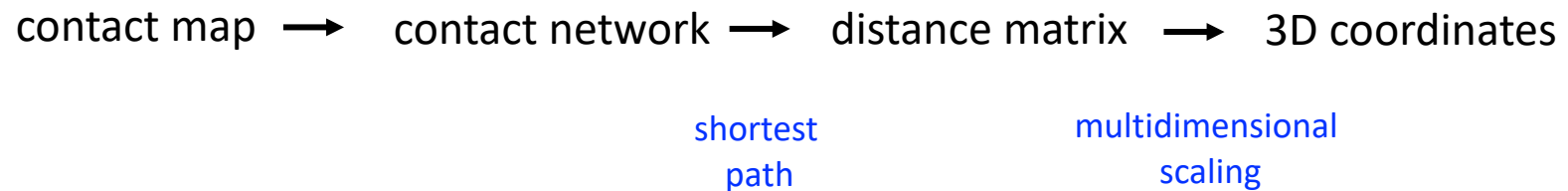


genome axis

The challenge: reconstruction of the underlying structure

- 3D protein structure
- **3D chromosome conformation**
- attractor: Hirata, Horai and Aihara EPJ 164:13 (2008)

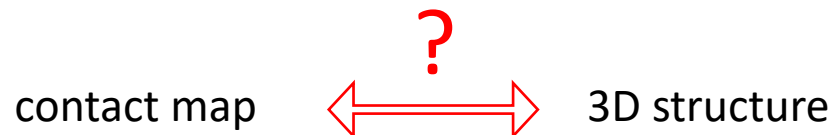
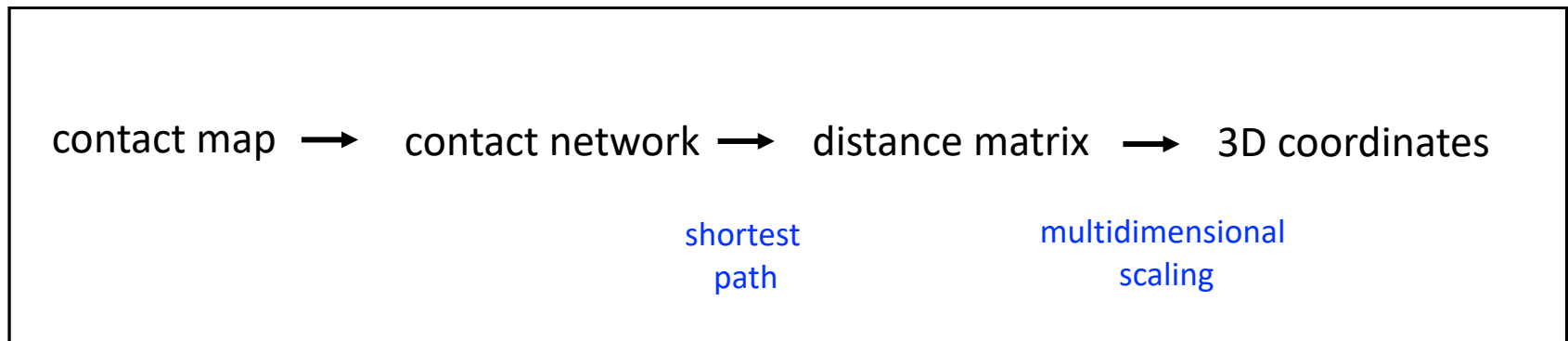
→ algorithm based on **shortest-path distance** and **multidimensional scaling**



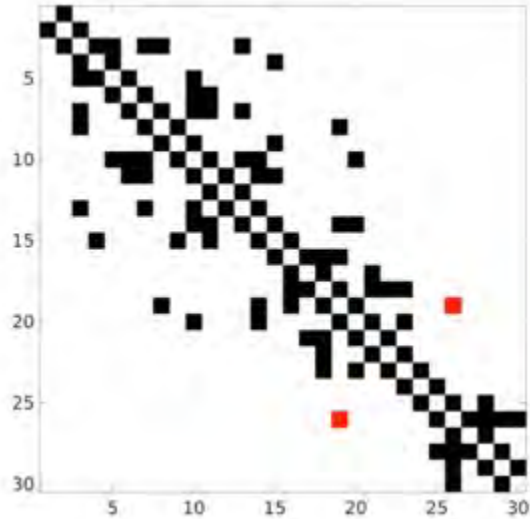
Another issue: reconstruction of the underlying structure

- 3D protein structure
- **3D chromosome conformation**
- attractor: Hirata and Aihara EPJ 164:13 (2008)

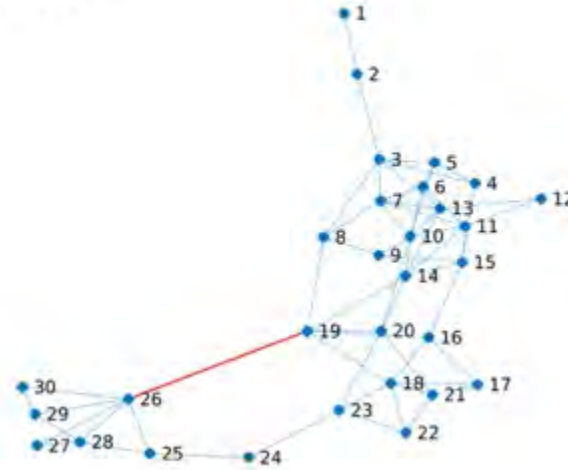
→ algorithm based on **shortest-path distance** and **multidimensional scaling**



From contact maps to contact networks



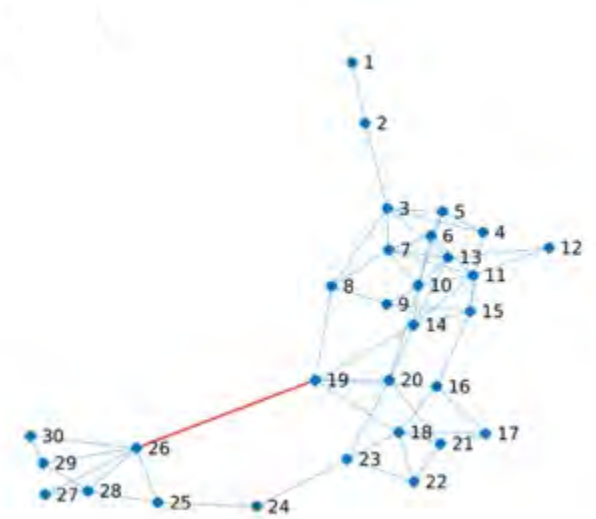
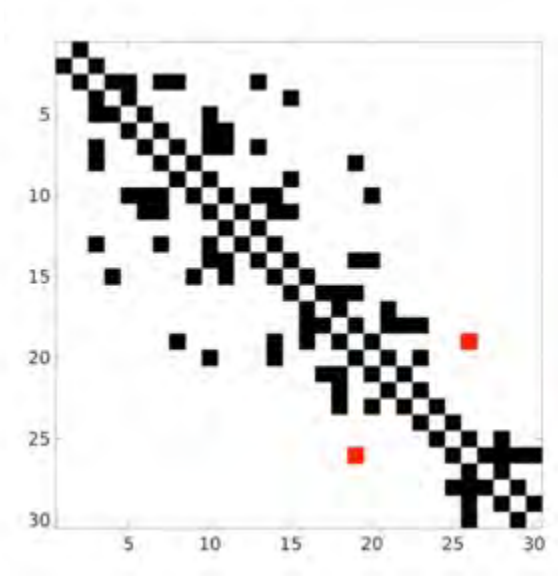
binary map A



edge $[i - j]$ if $A(i,j)=1$

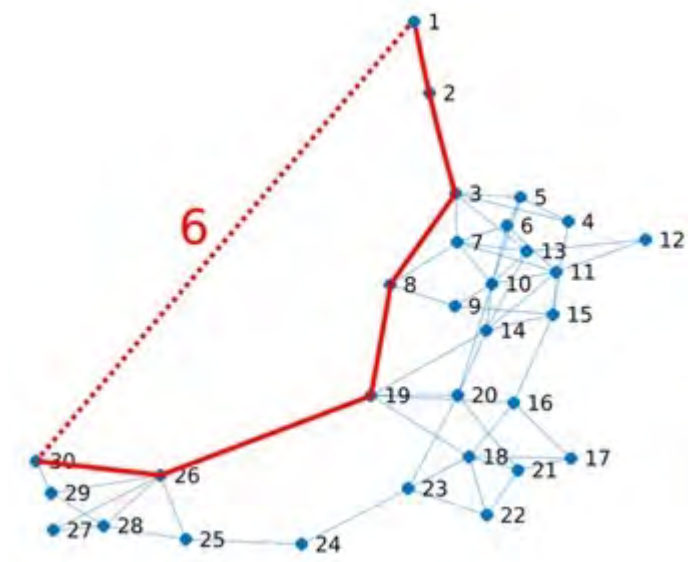
Morlot et al.
EPJ nbp 2016

From contact maps to contact networks to distances

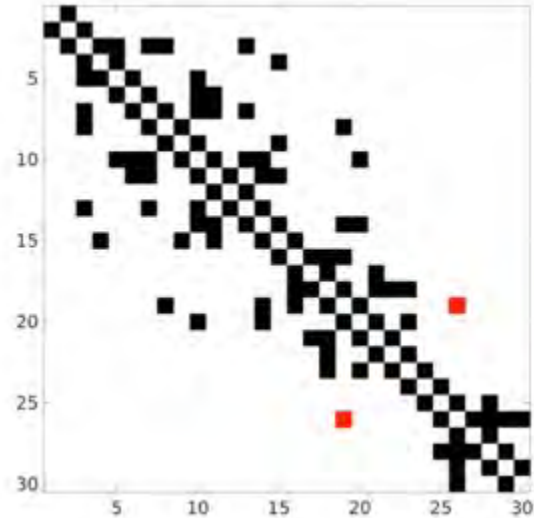


Distance = number of steps in the shortest path

Floyd –Warshall routing algorithm



From quantitative contact maps to weighted contact networks

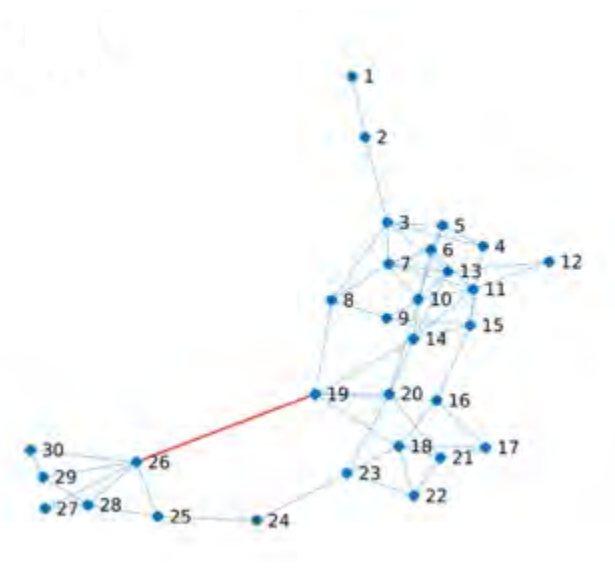
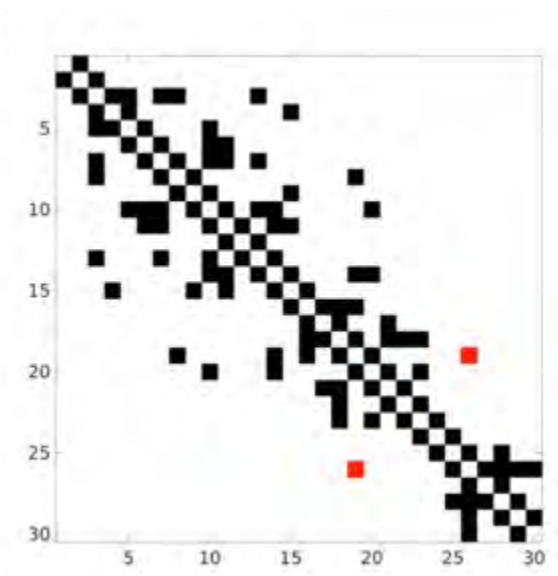


contact frequency $f(i,j)$



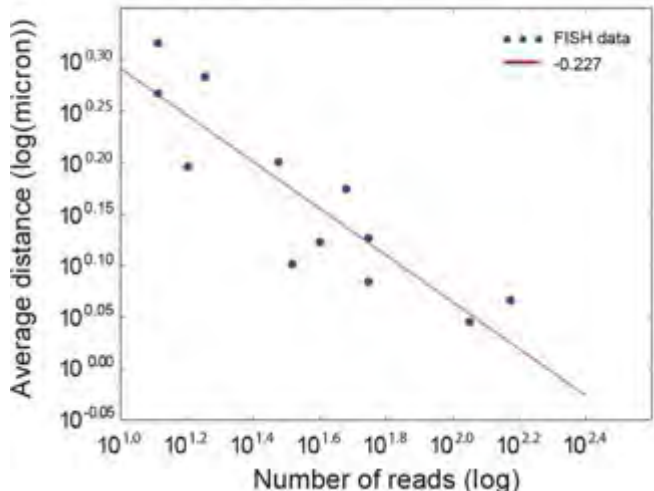
edge length $1 / f(i,j)^\alpha$

From contact networks to distances: A → D

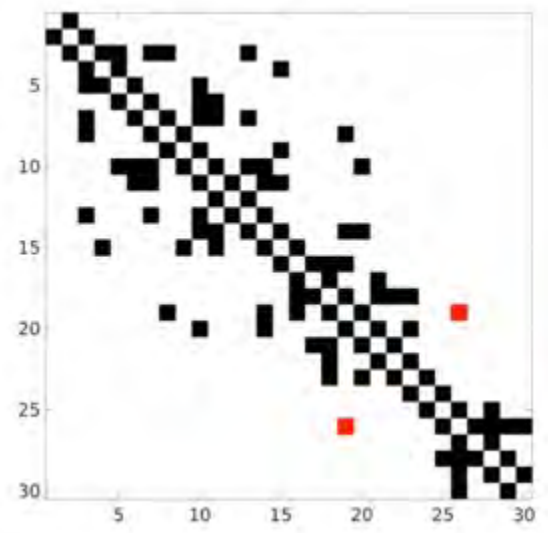


Comparison with FISH data: $\alpha = 0.2$

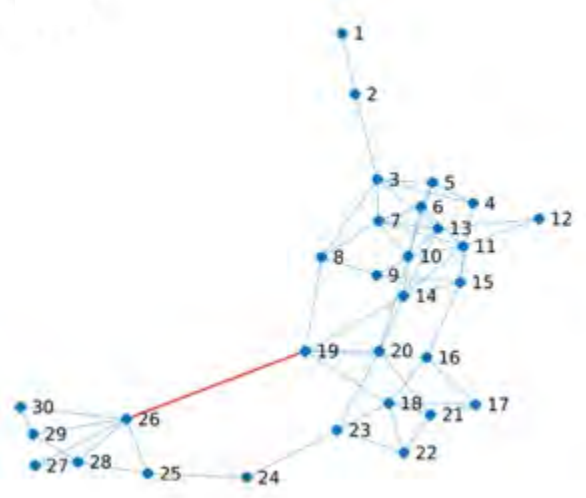
edge length $1 / f(i,j)^\alpha$



From contact networks to distances: $A \rightarrow D$



contact frequency $f(i,j)$



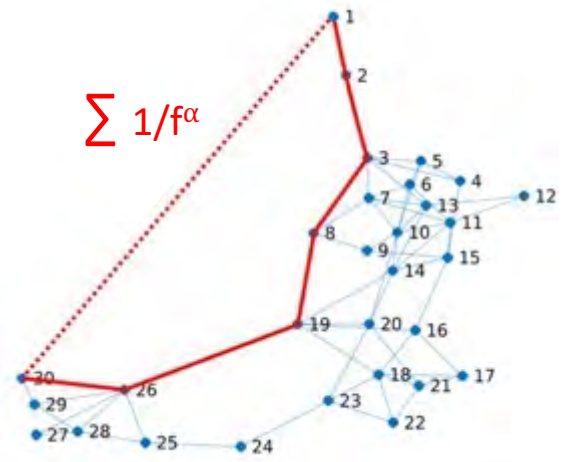
edge length $1 / f(i,j)^\alpha$

Morlot et al.
EPJ nbp 2016

Distance (i,j) = length of the shortest path

$$= \text{minimum over } \sum 1 / f(i,j)^\alpha$$

\rightarrow complete distance matrix D



Mathematical framework

Several matrices can be associated to a spatial structure (N points in 3D space)

Coordinate matrix **V** (points in 3D) $3 \times N$

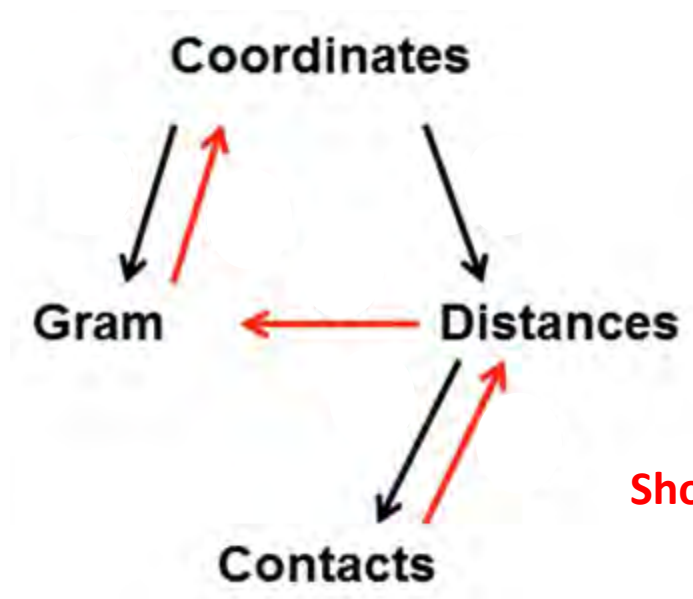
Gram matrix **G** (scalar products) $N \times N$

Distance matrix **D** (Euclidean distances) $N \times N$

Contact matrix **A** (binary or contact frequencies) $N \times N$

Distance geometry

Multi-dimensional scaling



Shortest-path distance

How to infer 3D coordinates from the knowledge of the distance matrix D

1) Solved long ago in mathematics: distance geometry (Havel et al. 1983)

- Distances D_{ij} → **metric matrix** $2M_{ij}=d_{oi}^2+d_{oj}^2-D_{ij}^2$ with d_{oi} expressed as a function of D_{kl}
- M has only 3 nonzero eigenvalues λ_a (condition: $x^T M x \geq 0$) and **M=G**
- **3D coordinates** are obtained from the 3 associated eigenvectors E_1, E_2, E_3 : $X_{i,a}=E_a(i) \lambda_a^{1/2}$
genomic loci $i=1\dots N$ and components $a=1,2,3$ a=1,2,3
i=1...N

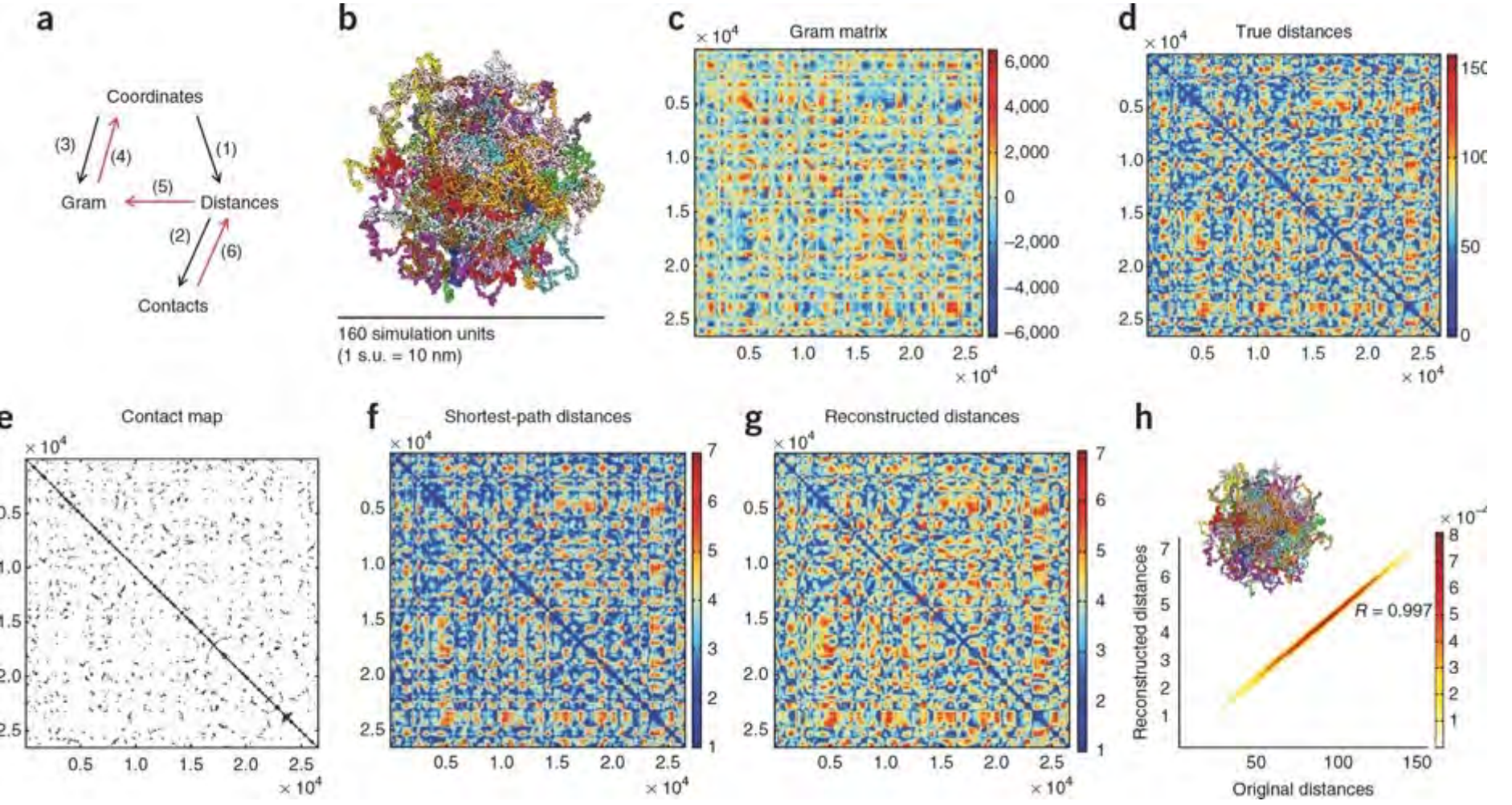
locus (X_1, X_2, X_3) with $X_1=E_1(\text{locus}) \lambda_1^{1/2}$, $X_2=E_2(\text{locus}) \lambda_2^{1/2}$ and $X_3=E_3(\text{locus}) \lambda_3^{1/2}$

2) In practice: multidimensional scaling

- M has more than 3 nonzero eigenvalues → **truncation** by keeping the first 3 eigenvalues = **G**
- Truncated matrix G corresponds to the **best approximating 3D structure**
- **3D coordinates** are obtained from the 3 associated eigenvectors as above

→ **structure up to a rotation, symmetry and dilation**

Simulated data as a benchmark



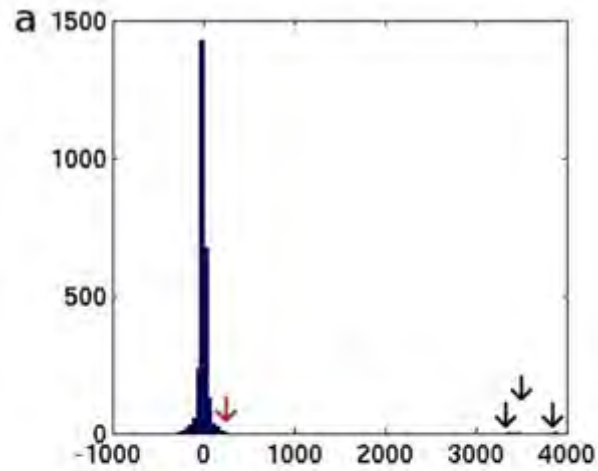
b: simulated yeast nucleus (polymer model) \rightarrow **h:** reconstructed 3D structure

Validation with synthetic data

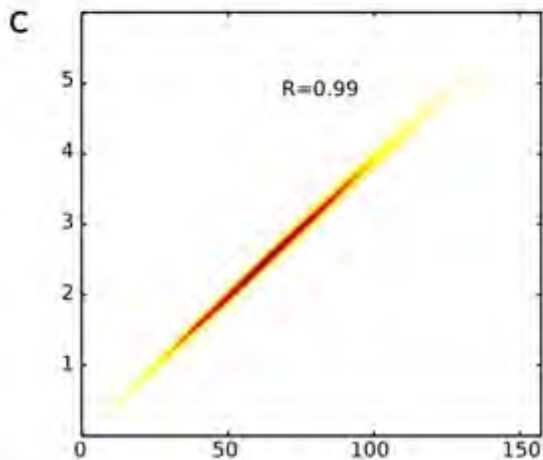
simulated structure(s) → contact map → 3D structure

using shortest-path distance

Spectral validation
(metric matrix M)



Spearman rank
correlation coeff. R
(original vs reconstructed)

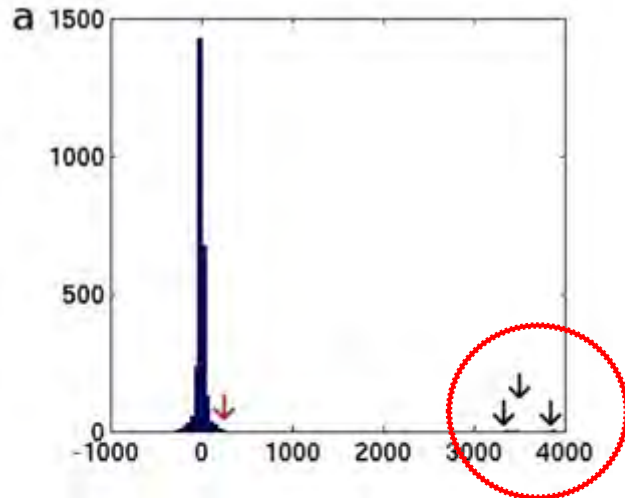


Validation with synthetic data

simulated structure(s) → contact map → 3D structure

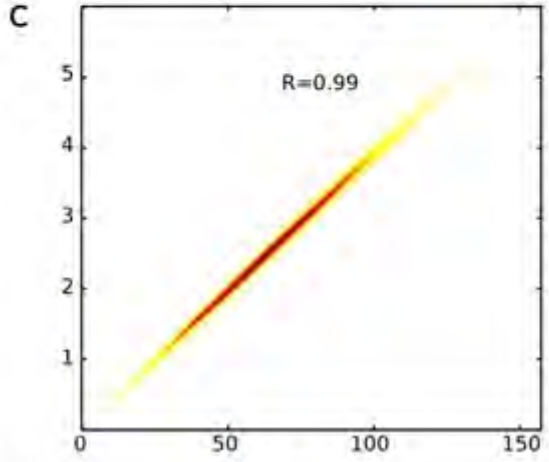
using shortest-path distance

Spectral validation
(matrix M)



3D truncation
→ Gram matrix G
of a 3D structure

Spearman rank
correlation coeff. R
(original vs reconstructed)



3D reconstruction at different scales

a: unit= 1 restriction fragment (about 3kb)

b: unit= 50 restriction fragments

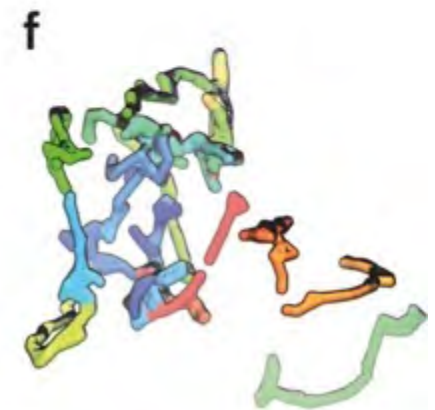
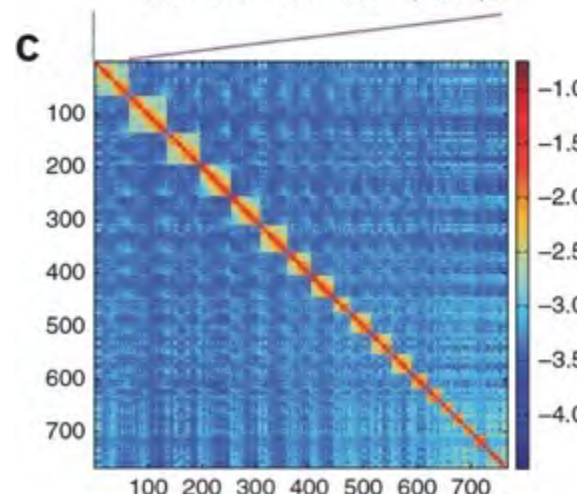
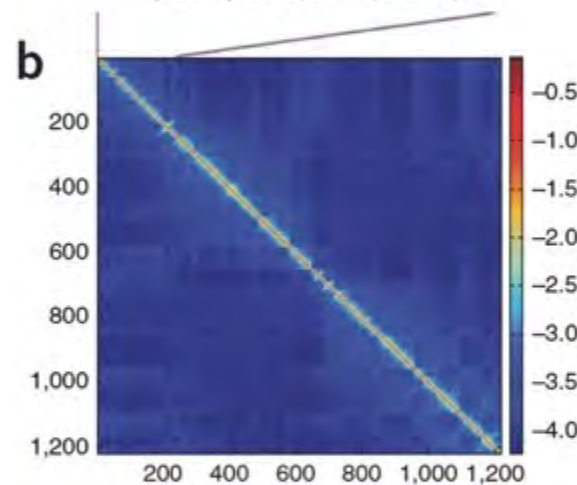
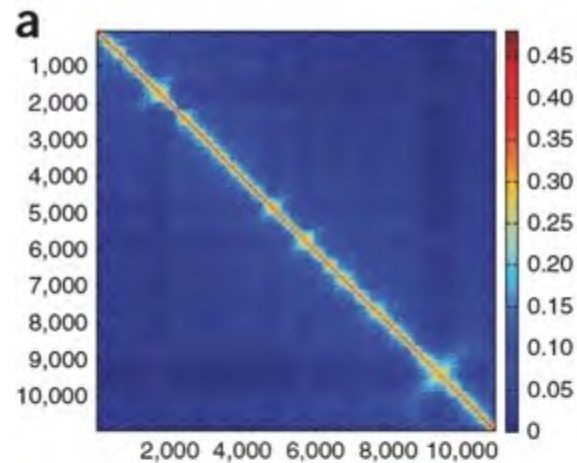
c: unit= 1000 restriction fragments

a-d: part of Human chromosome 1

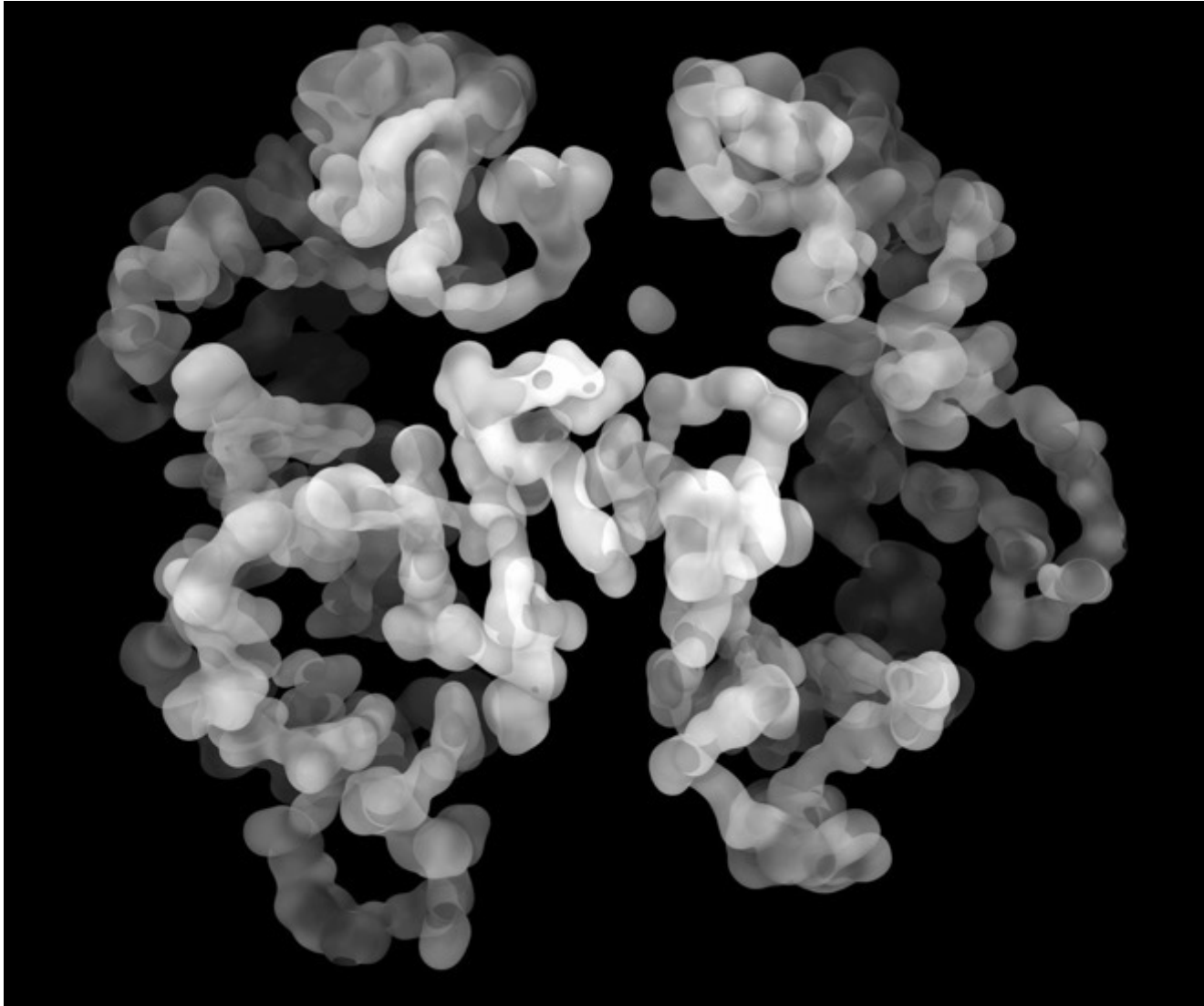
b-e: whole about chromosome 1

c-f: all human autosomal chromosomes

(human embryonic stem cells; data from Dixon et al. 2012 and Kalhor et al. 2012)



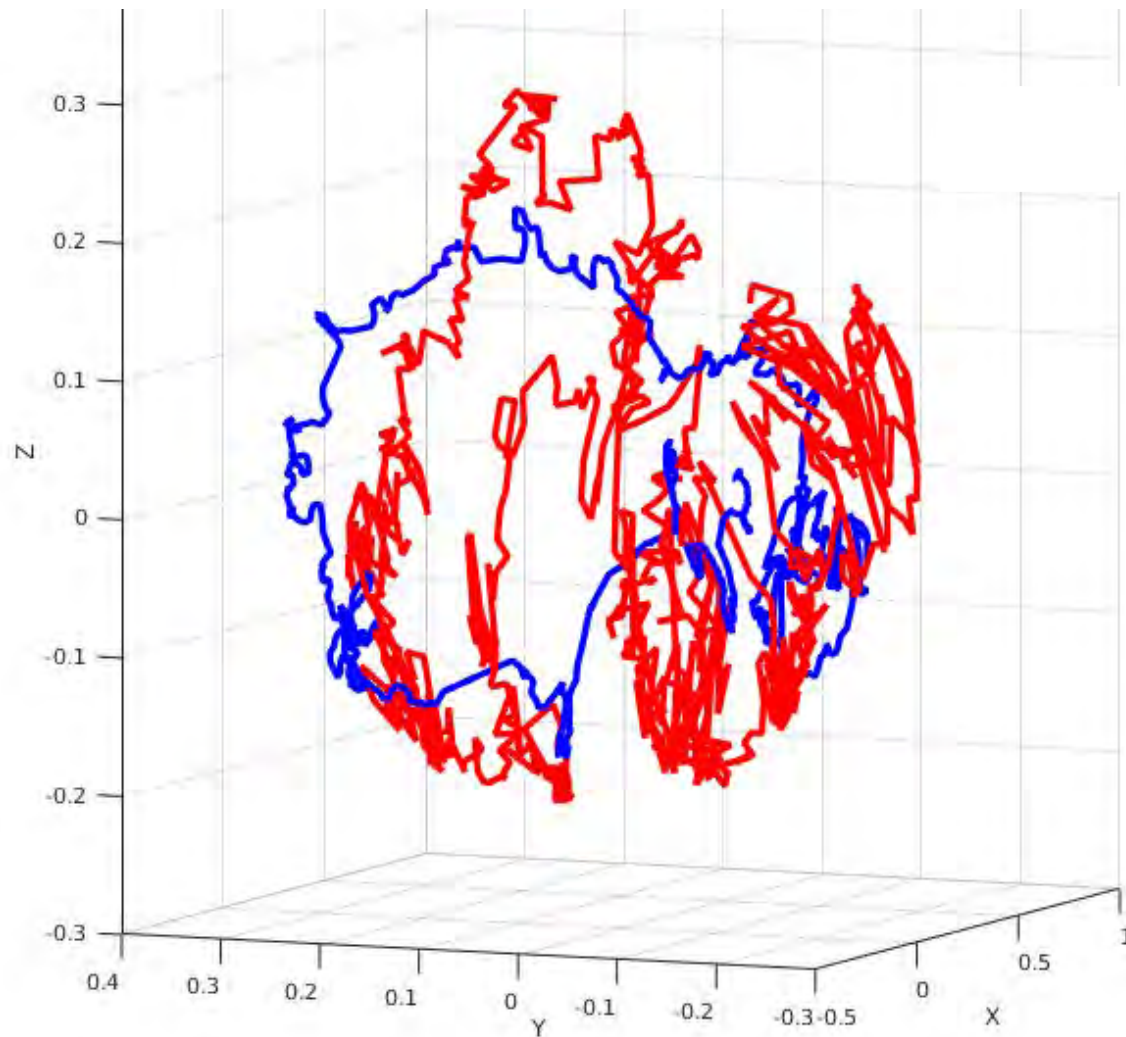
Art exhibition 'Le Rêve des Formes', Palais de Tokyo, Paris 2017



Human chr1

Methodological improvements: tuning the exponent α

edges of length $1/f_{ij}^\alpha$ and α can be tuned



$\alpha=0.2$: local accuracy
(consistent with FISH data)

$\alpha=1$: global shape

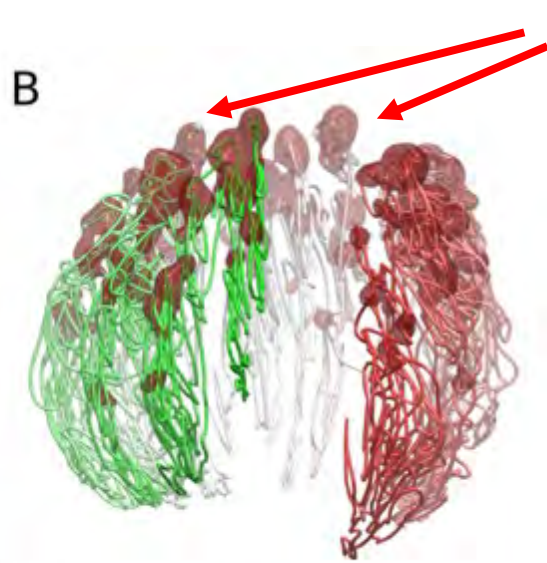
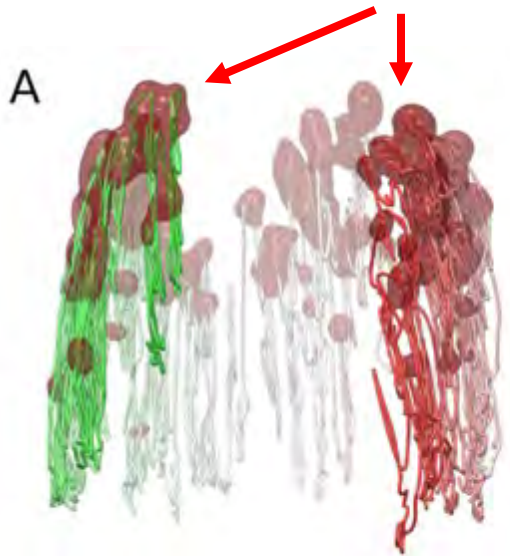
Various options for multi-dimensional scaling (MDS)

Human chromosome 2
GM12878 cells

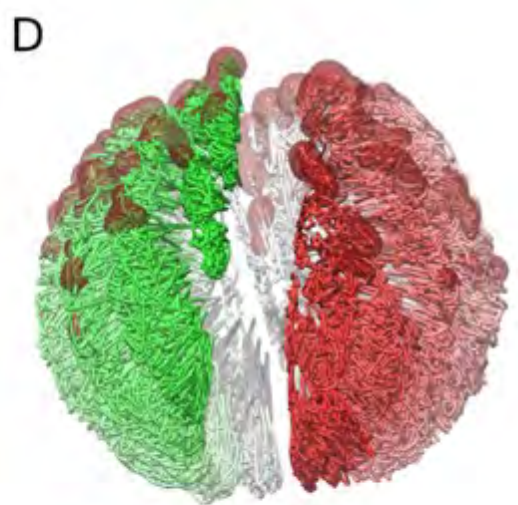
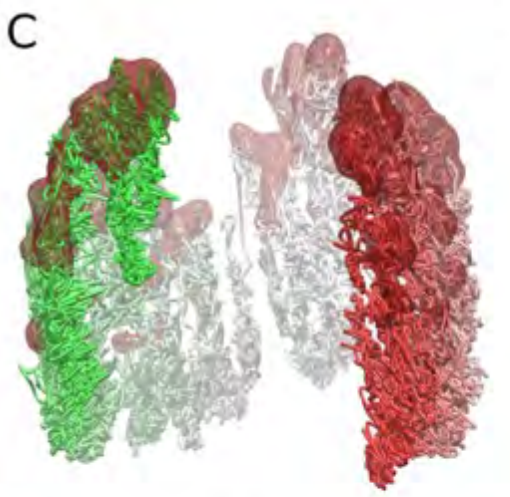
strain-based MDS

Sammon's MDS

resolution
100kb



resolution
10kb



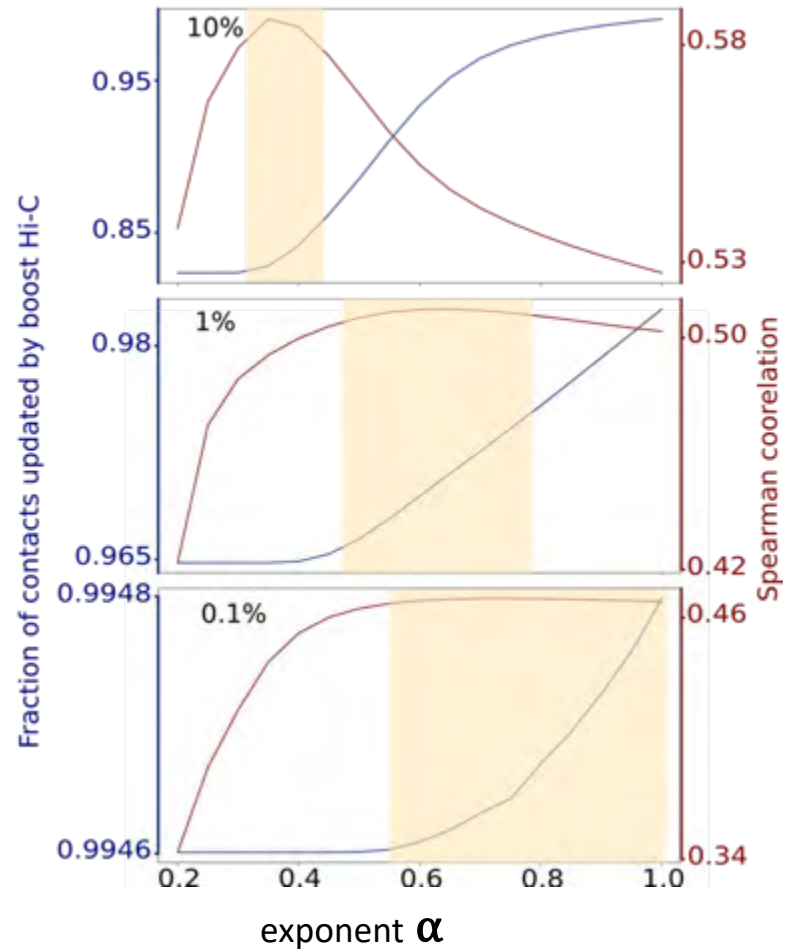
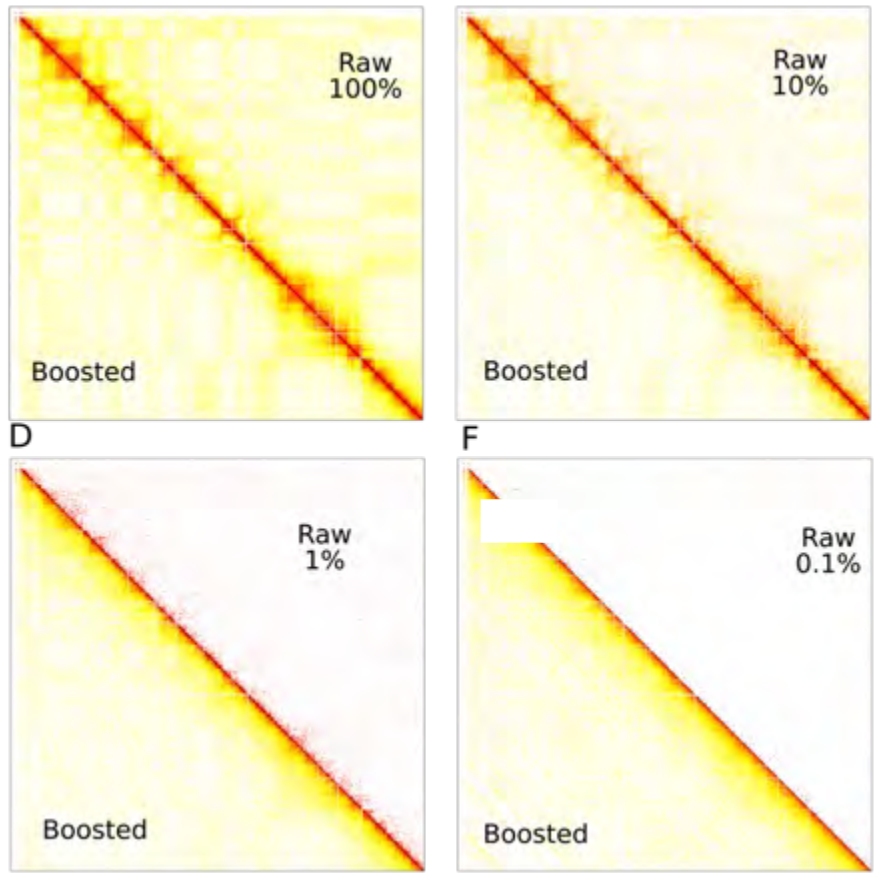
Carron et al. MiMB 2021

0 chr2 240 Mbp

Active promoter region

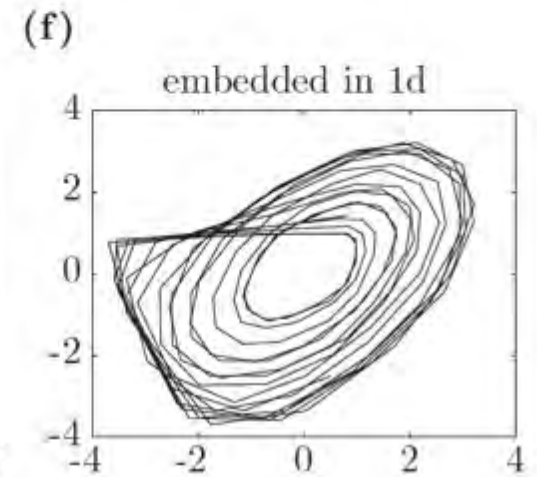
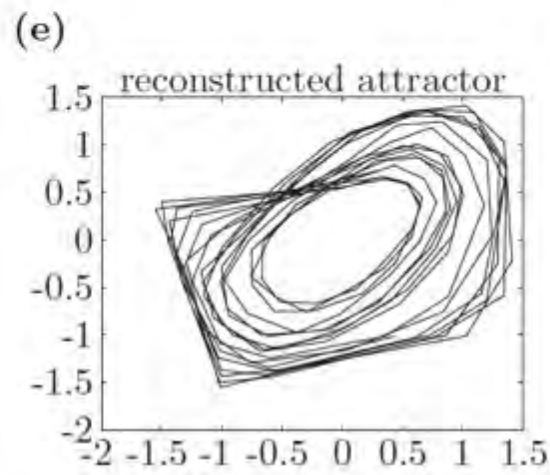
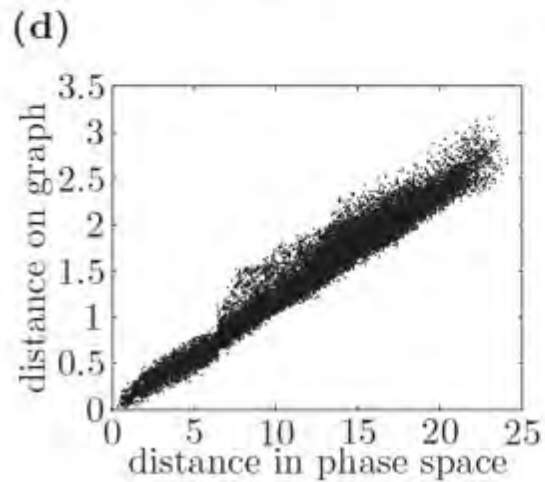
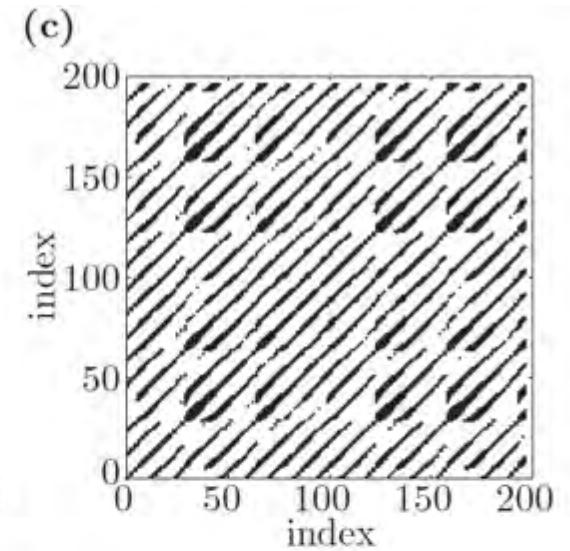
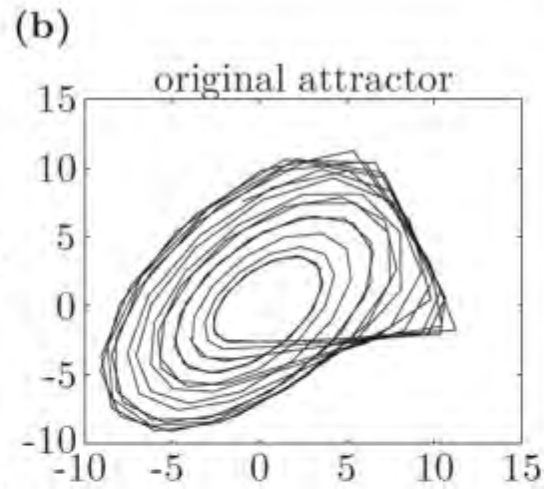
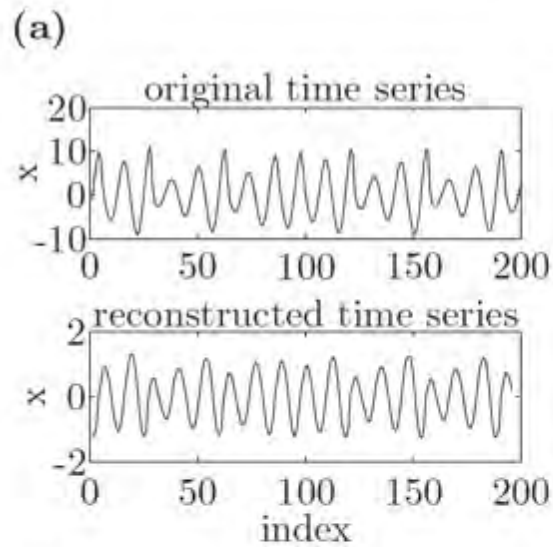
Shortest-path distance for Hi-C map enhancement (*Boost-HiC*) -- missing data completion

Hi-C map $f^0(i,j)$ \rightarrow weighted contact network \rightarrow shortest-path distance $d(i,j)$ \rightarrow improved contact freq.
 edge length $1/f^0(i,j)^\alpha$ for any pair (i,j) $f(i,j) = 1/d(i,j)^\alpha$



Best when only vanishing/small components are updated

In the RP context: Hirata, Horai and Aihara, EPJ Special Topics 164:13 (2008)



reconstruction of an attractor:

RP= **binary** contact map

contact network with edges of different lengths

Jaccard distance between the neighborhoods of i and j

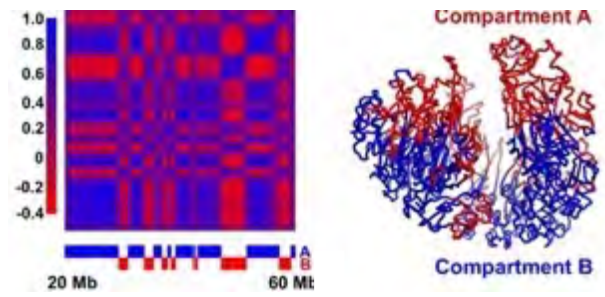
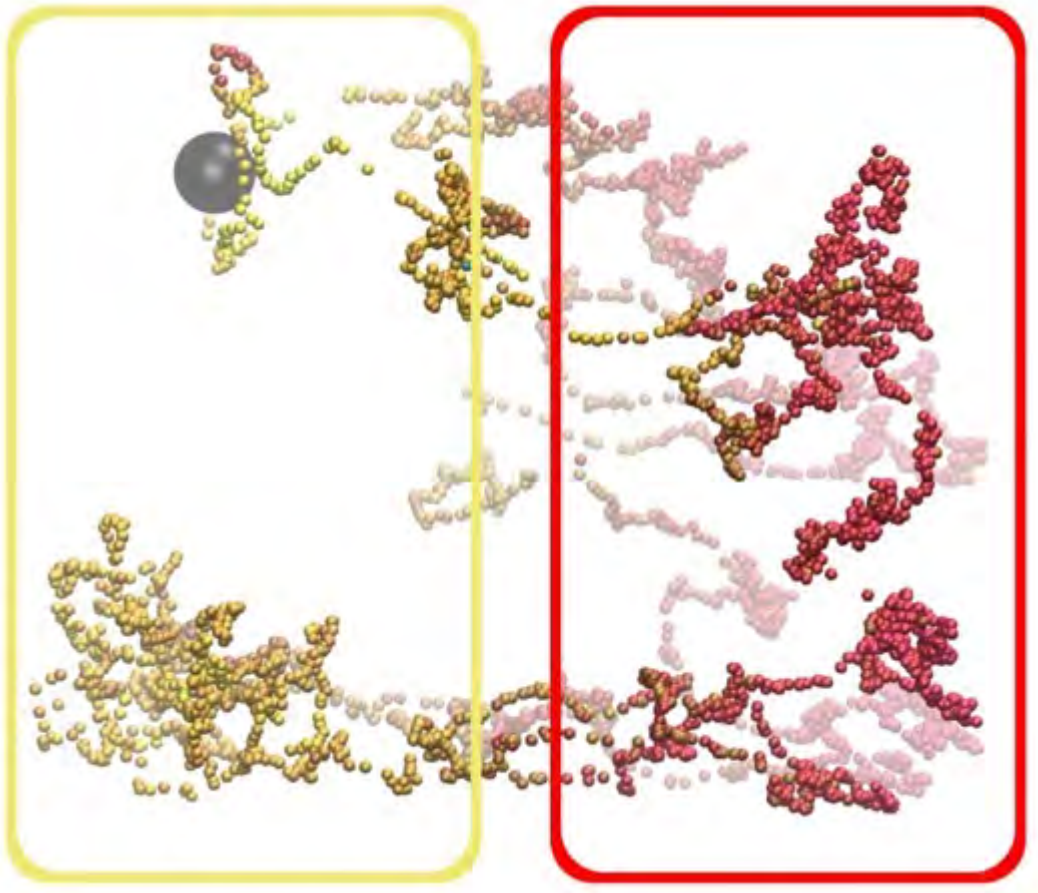
length $w(i, j) = 1 - \frac{|G_i \cap G_j|}{|G_i \cup G_j|}$

→ relies on the network topology only
specially well-suited for binary contact maps

→ efficient for single-cell HiC (Hirata, Oda, Ohta & Aihara, Scientific Reports 2018)

Results: chromosome compartments

A/B compartments



principal eigenvector
of the correlation matrix

Human chromosome 1

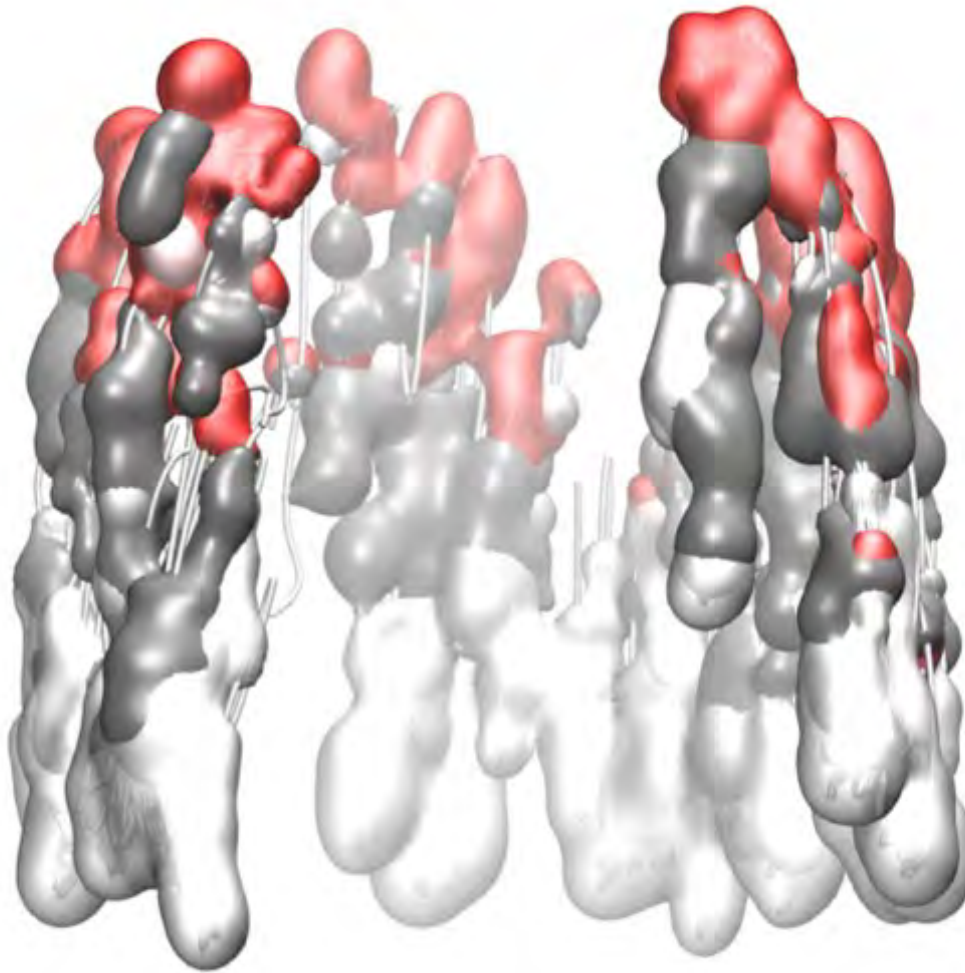
Lesne et al. Nat. Meth. 2014

Yellow: gene-rich, GC-rich = active compartment A
Red: gene-poor, AT-rich = inactive compartment B

Results: chromosome compartments

3 epigenetic states: **Active**
Bivalent (Polycomb-rich)
Quiescent

Human chromosome 2

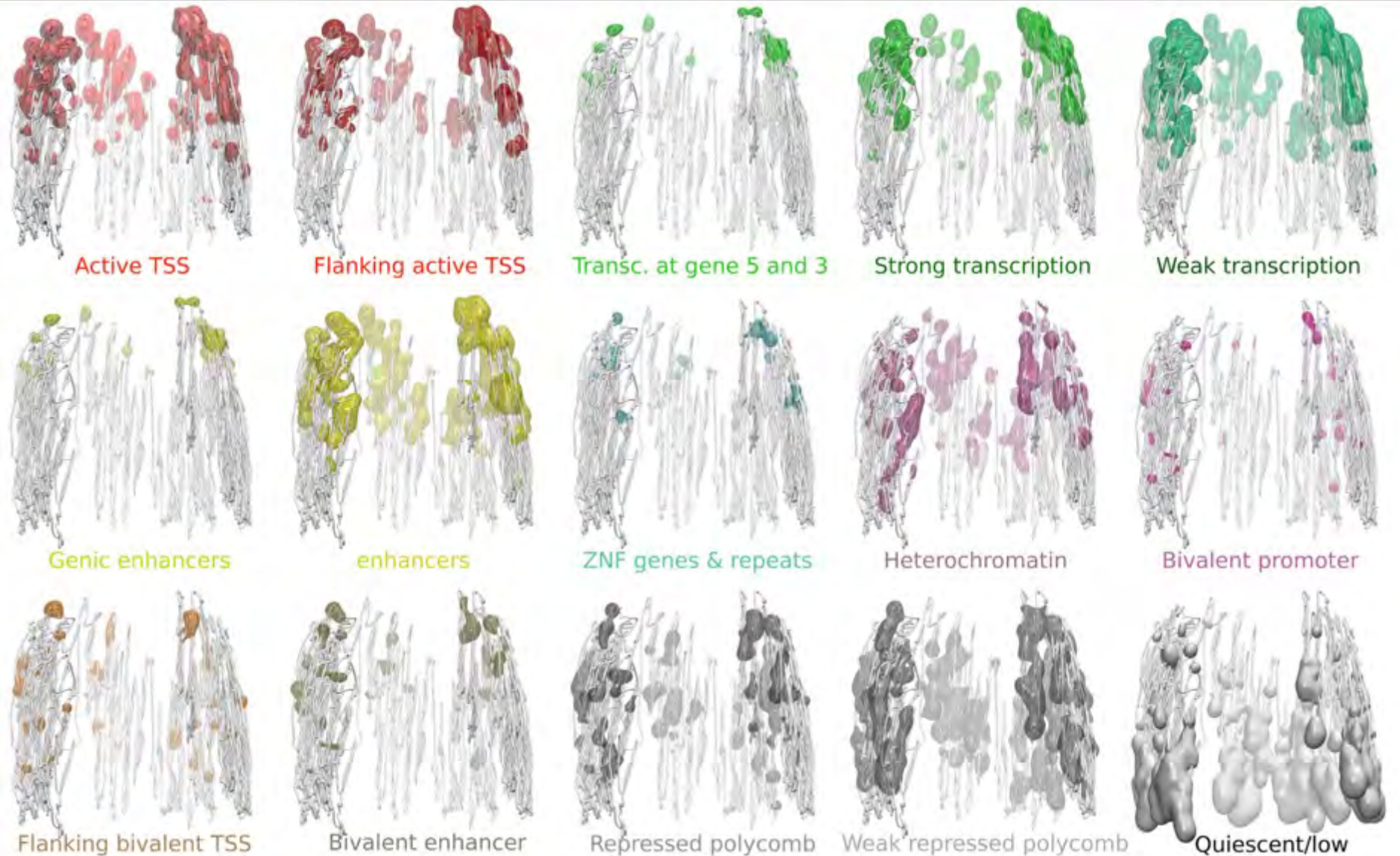


Carron et al. MiMB 2021

→ **3 spatial compartments:** active A (red), inactive B (white), and an intermediary one (grey)

3D genome browser: linear annotations in their 3D context

Annotating genomic loci with **15 epigenetic states**

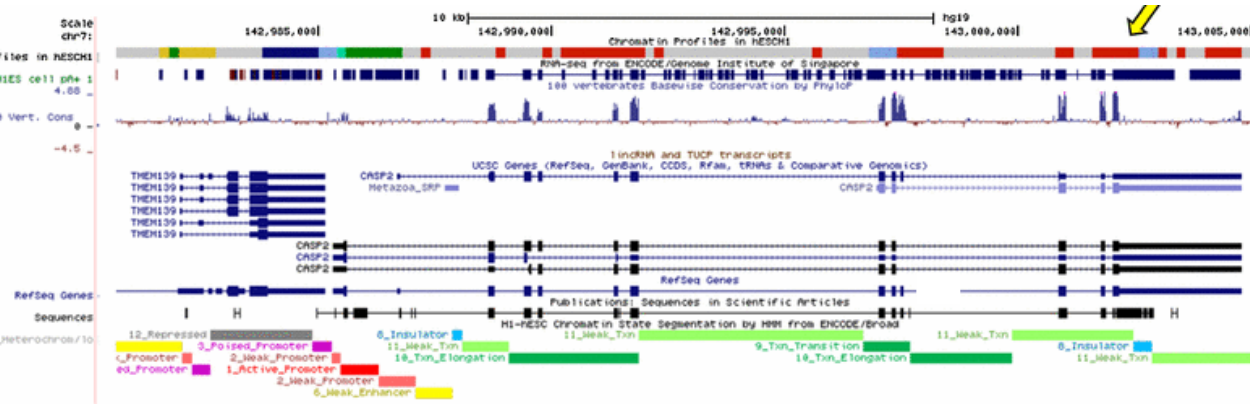


human chromosome 2

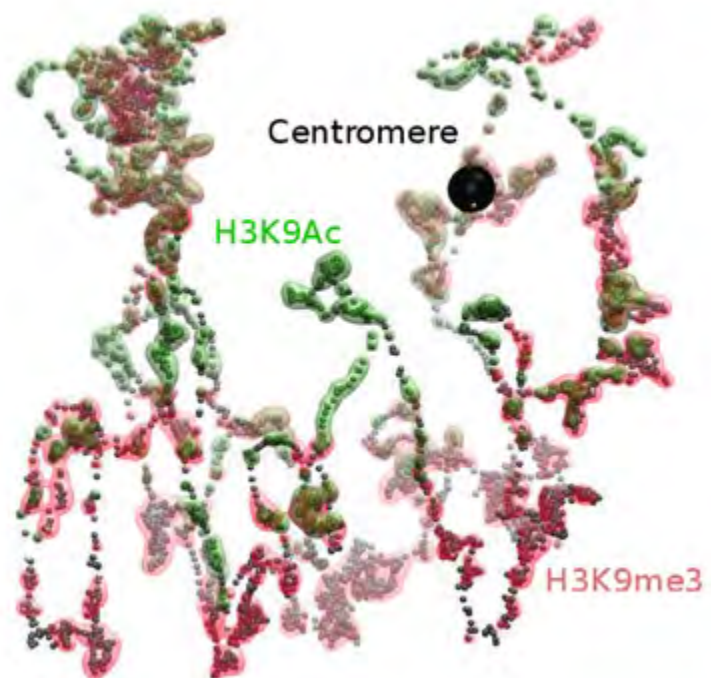
Carron et al. MiMB 2021

Dual view: linear sequences vs 3D structure

→ 1D genome browser



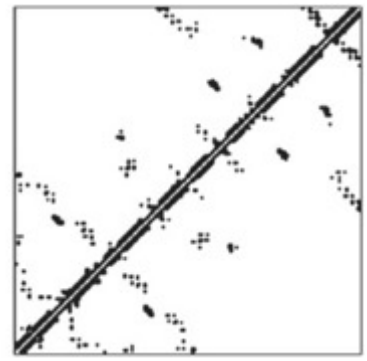
→ 3D genome browser



Contact maps

$A_{ij} = 1$ means that i and j are in contact (or close in space)

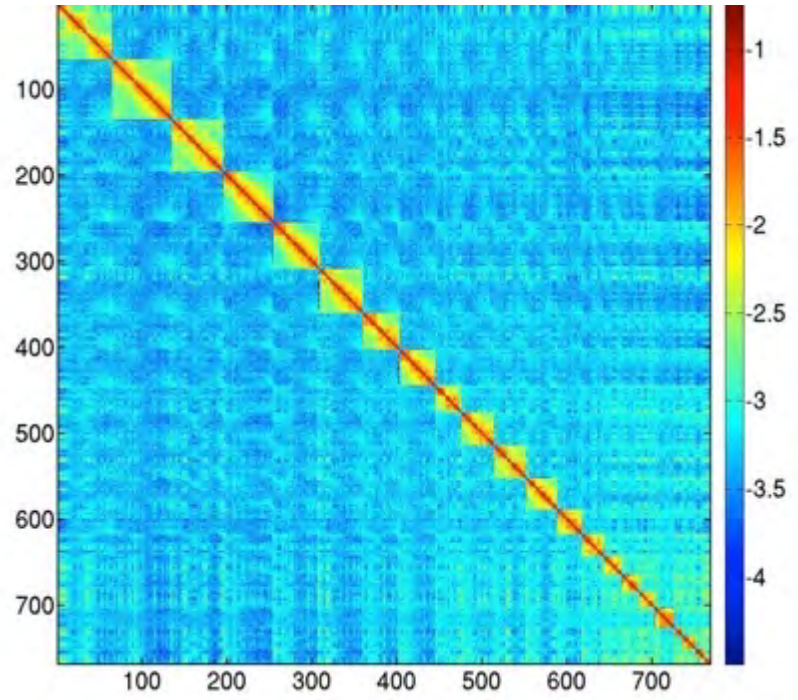
Protein structure



Colosimo & Giuliani 2015

real space
axes: residues

Chromosome conformation

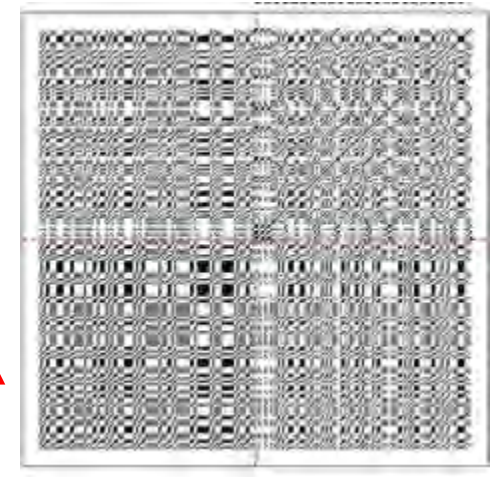


Lesne et al. 2014

Red = high contact frequency

real space
axes: position along the genome
 f_{ij} = contact frequency

Recurrence plots



Eckmann, Kamphorst & Ruelle EPL 1987

Marwan et al. Physics Reports 2007

phase space
axes: time

(original paper)

3D genome reconstruction from chromosomal contacts

A. Lesne, J. Riposo, P. Roger, A. Cournac, J. Mozziconacci

Nature Methods 11, 1141–1143 (2014)

<https://hal.archives-ouvertes.fr/hal-02324120>

Network concepts for analyzing 3D genome structure from chromosomal contact maps

J.B. Morlot, J. Mozziconacci, A. Lesne

European Physical Journal: Nonlinear Biomedical Physics 4, 2 (2016)

<https://link.springer.com/article/10.1140/epjnbp/s40366-016-0029-5>

(tutorial)

The 3D organization of chromatin colors in mammalian nuclei

L. Carron, J.B. Morlot, A. Lesne, J. Mozziconacci

Methods in Molecular Biology, vol. 2301, Springer 2021

<https://hal.archives-ouvertes.fr/hal-03172656>

Boost-Hi-C: computational enhancement of chromosomal contact map resolution

L. Carron, V. Matthys, J.B. Morlot, A. Lesne, J. Mozziconacci

Bioinformatics 35, 2724-2729 (2019) hal-02349503 bioRxiv, 471607

<https://hal.archives-ouvertes.fr/hal-02349503>

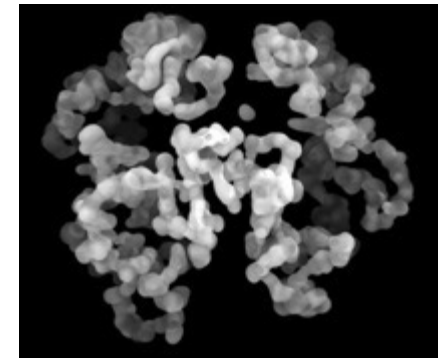


THANK YOU FOR YOUR ATTENTION

and special thanks to my collaborators:

Julien MOZZICONACCI (MNHN, Paris)

Sofiane RAMDANI (University of Montpellier, LIRMM)



Tutorial paper for ShRec3D: The 3D organization of chromatin colors in mammalian nuclei.
L. Carron, J.B. Morlot, A. Lesne, J. Mozziconacci
Methods in Molecular Biology, vol. 2301, Springer 2021
<https://hal.archives-ouvertes.fr/hal-03172656>

