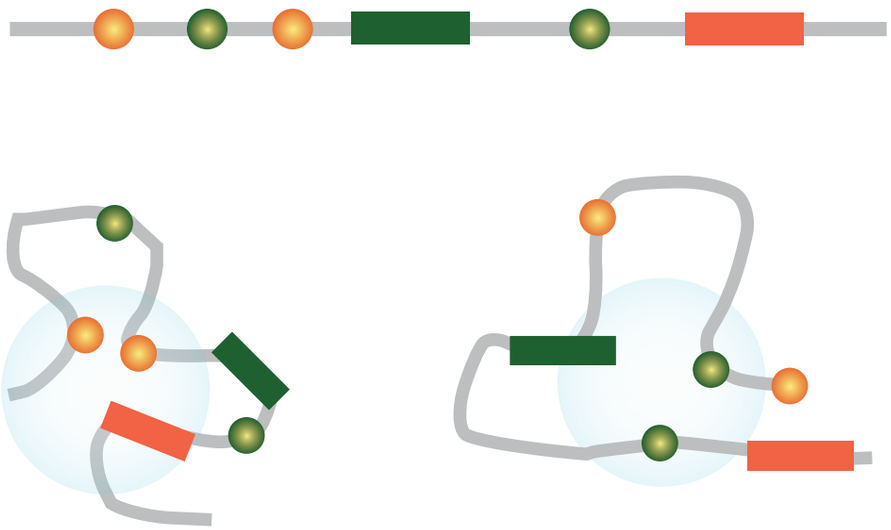
Does color have structure?

Modeling 3D domains of the fly genome.

Marc A. Marti-Renom
Genome Biology Group (CNAG)
Structural Genomics Group (CRG)

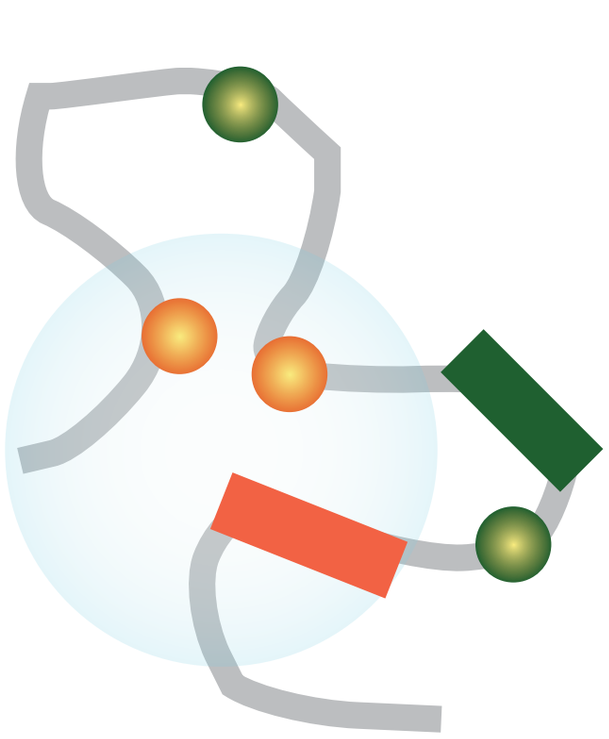
*iCrea
INSTITUCIÓ CATALANA DE RECERCA I ESTUDIS AVANÇATS

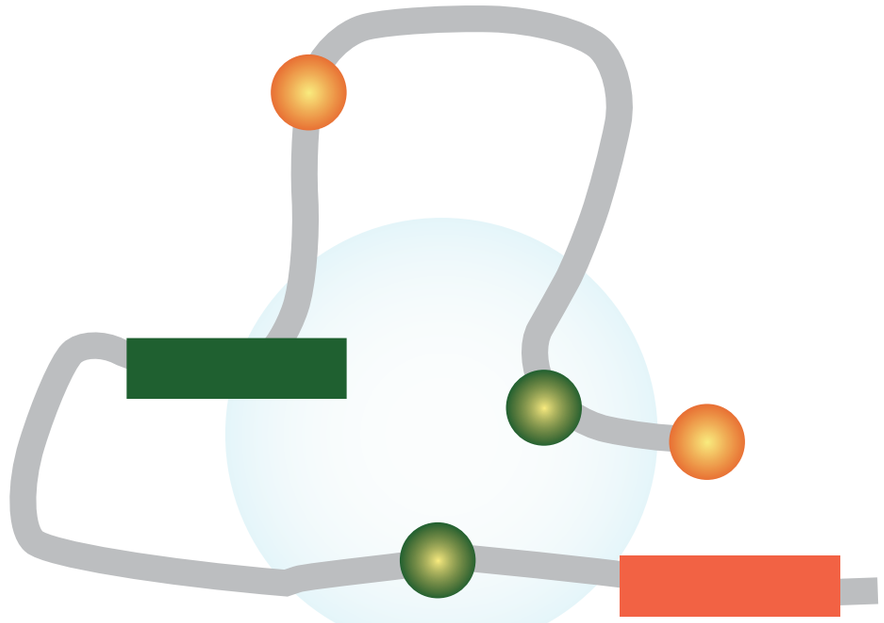
cnag CRG Centre for Genomic Regulation



TADbit TADkit

cnag CRG Centre for Genomic Regulation

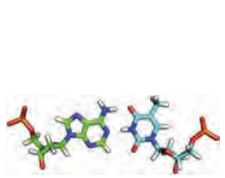


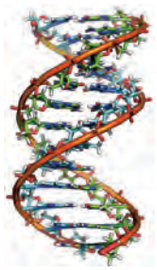


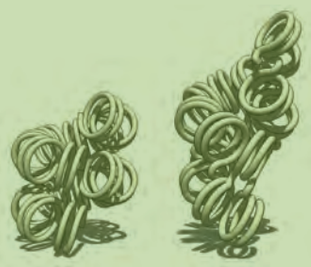
Resolution Gap

Marti-Renom, M. A. & Mirny, L. A. PLoS Comput Biol 7, e1002125 (2011)

Knowledge

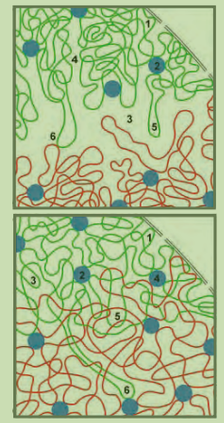


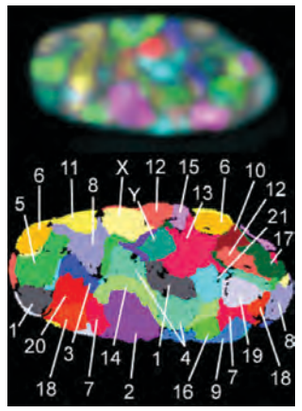




IDM

INM





DNA length

10^0 10^3 10^6 10^9 nt

Volume

10^{-9} 10^{-6} 10^{-3} 10^0 10^3 μm^3

Time

10^{-10} 10^{-8} 10^{-6} 10^{-4} 10^{-2} 10^0 10^2 10^3 s

Resolution

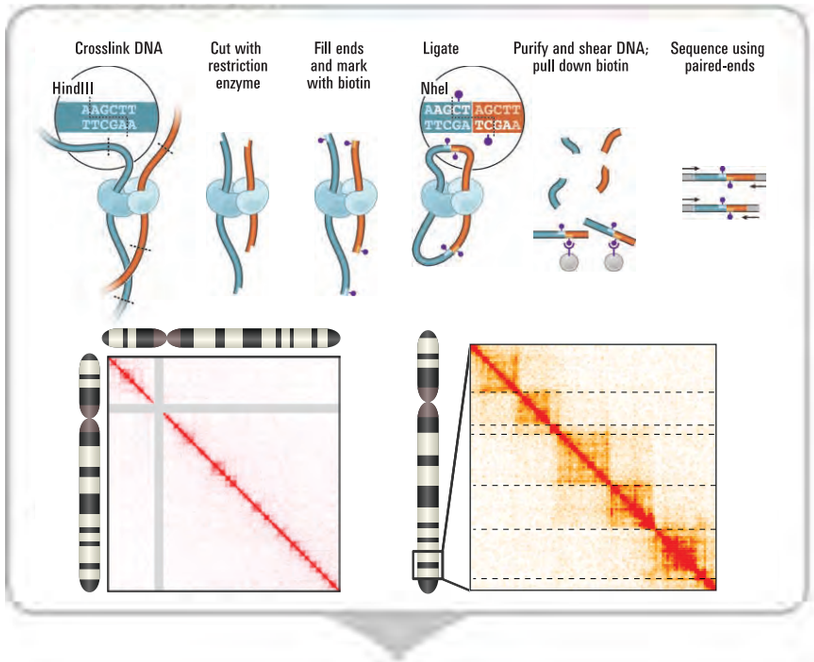
10^{-3} 10^{-2} 10^{-1} μ

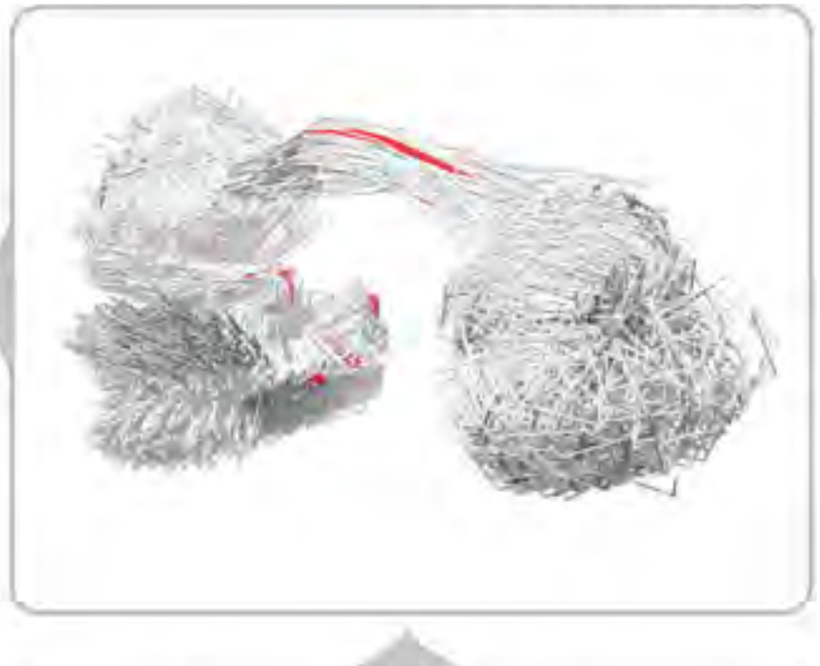
cnag CRG Centre for Genomic Regulation

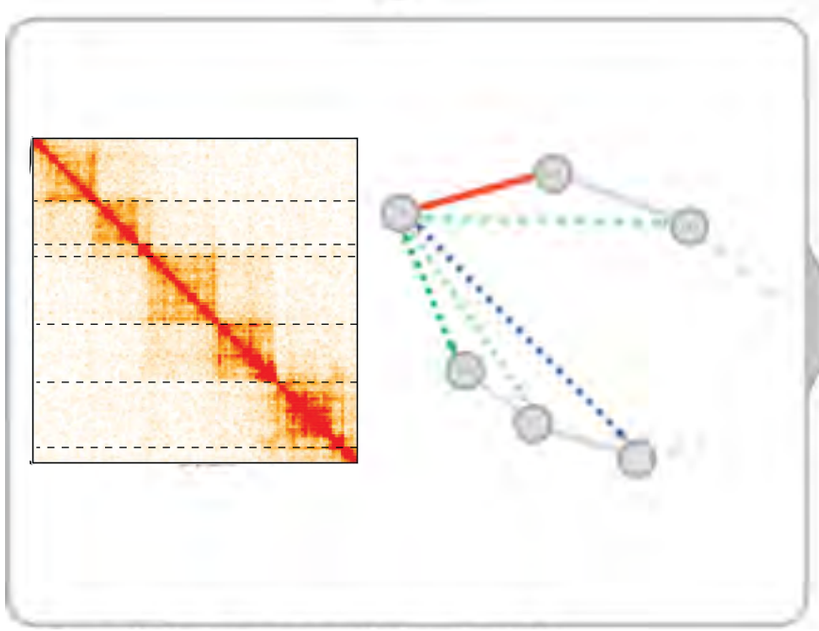
Hybrid Method

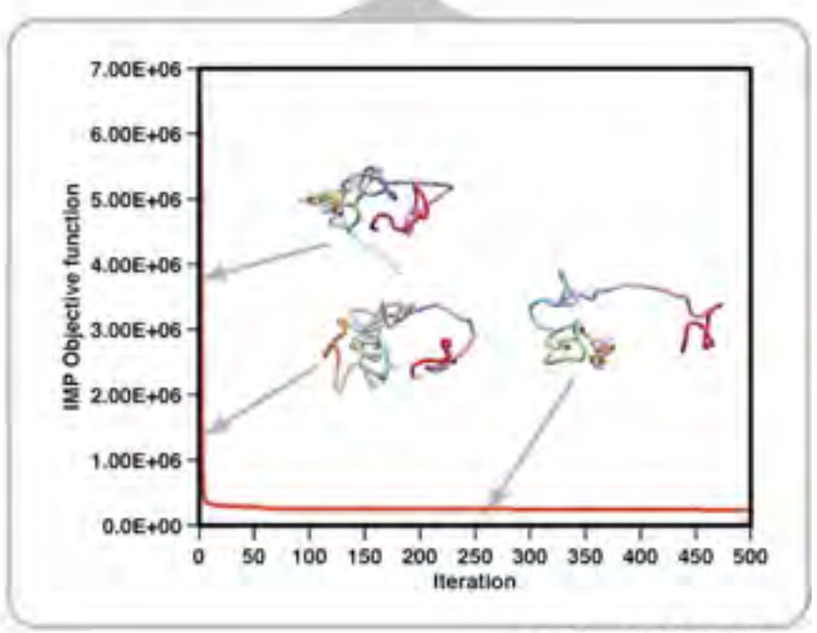
Baù, D. & Marti-Renom, M. A. Methods 58, 300–306 (2012).

Experiments

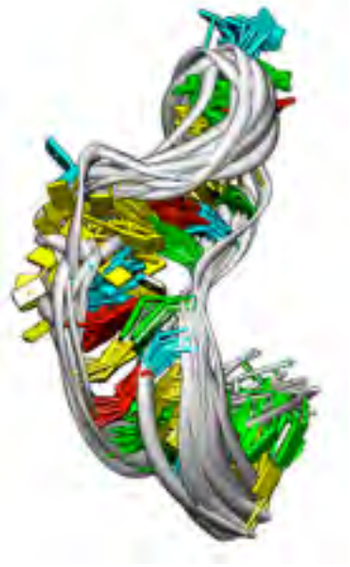


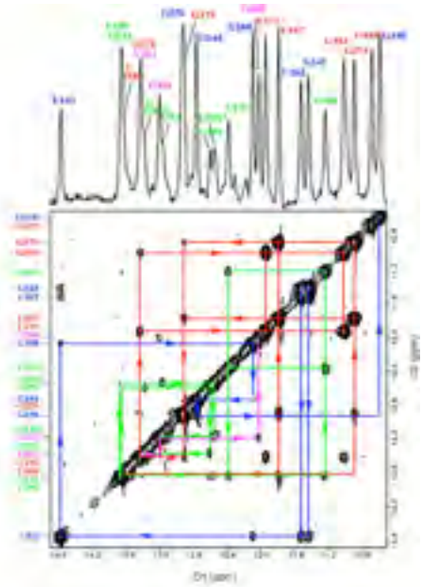






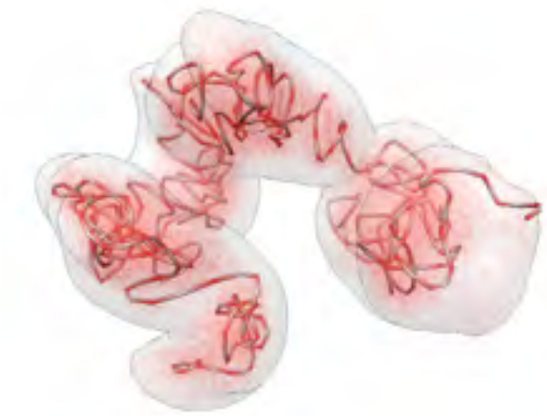
Computation

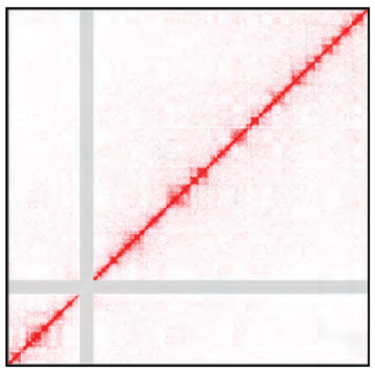




Biomolecular structure determination

2D-NOESY data



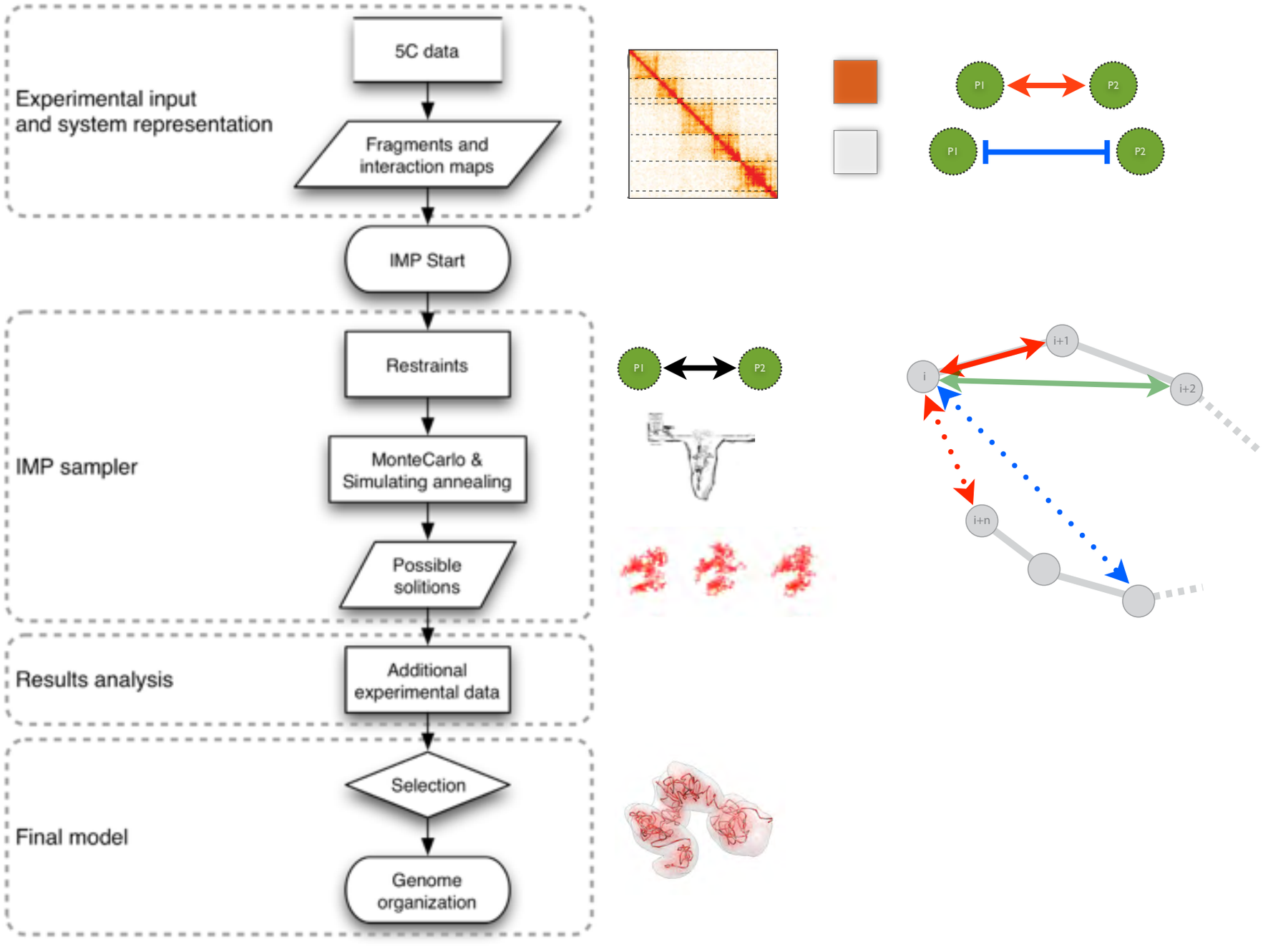


Chromosome structure determination

3C-based data

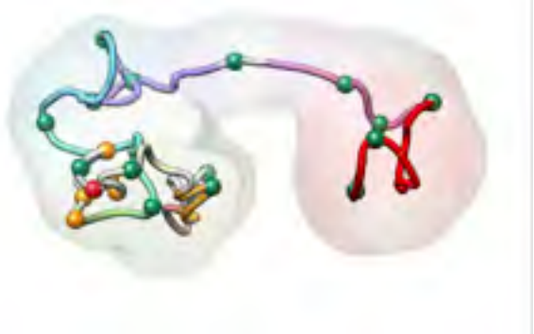
TADbit

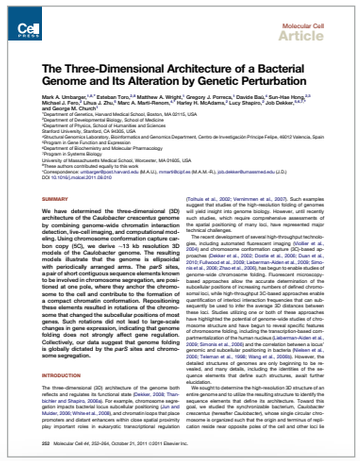
http://3DGenomes.org

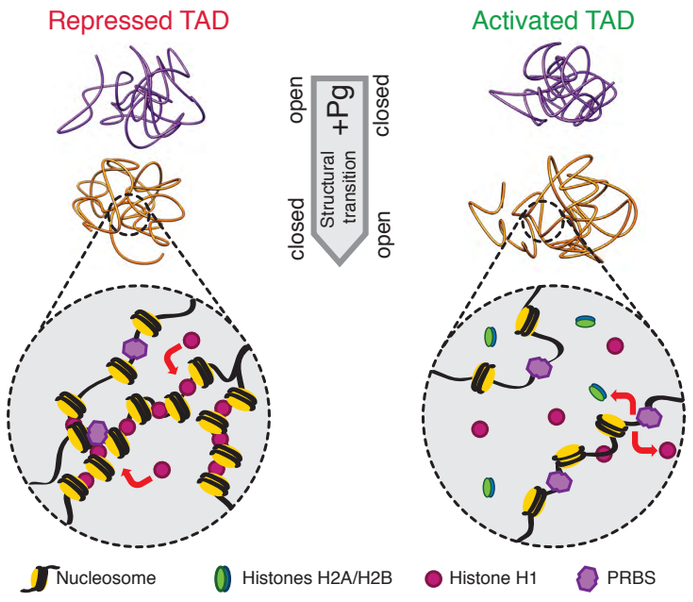


TADbit previous applications....

Baù, D. et al. Nat Struct Mol Biol (2011).
Umbarger, M. A. et al. Mol Cell (2011).







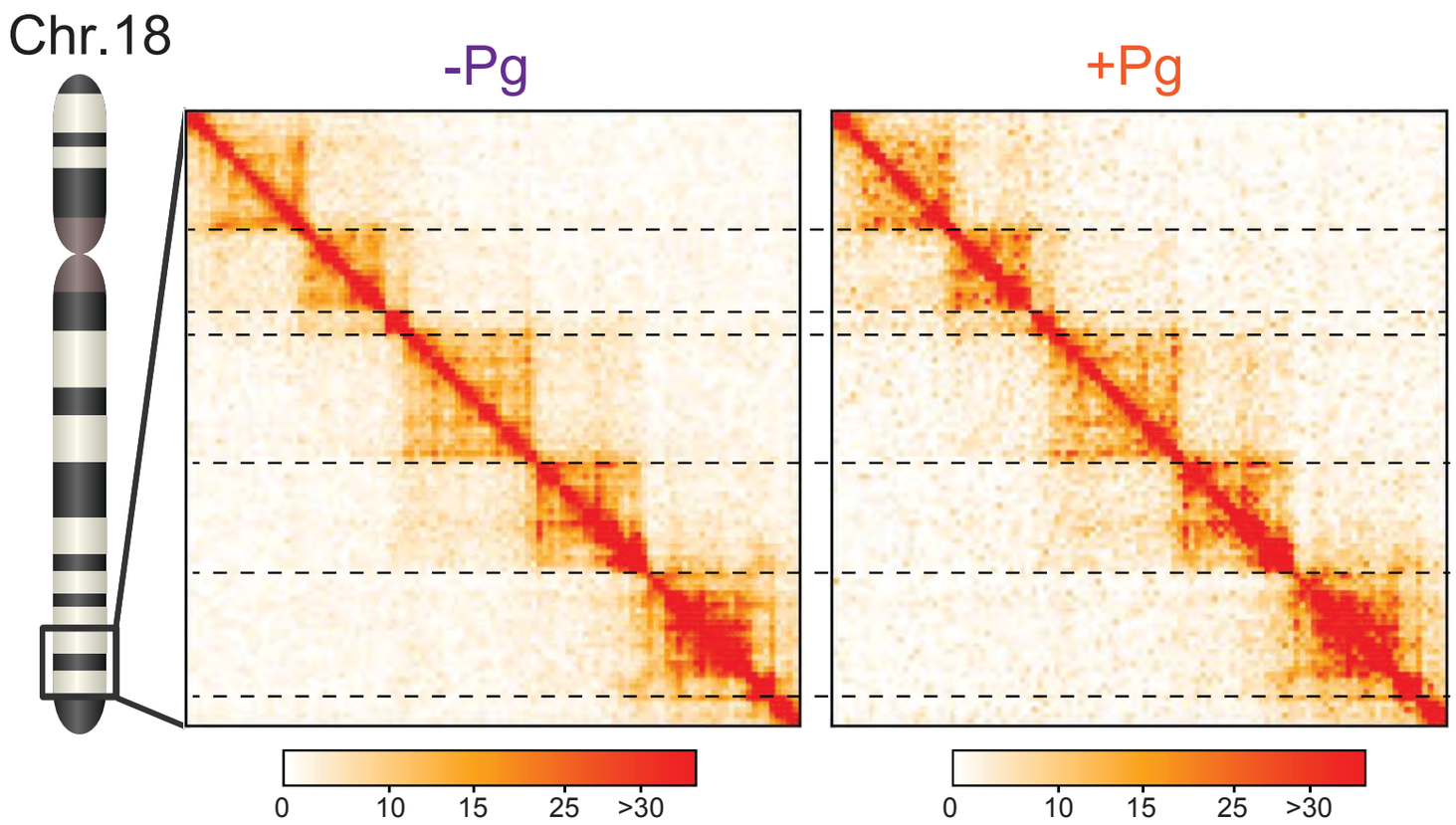
Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation

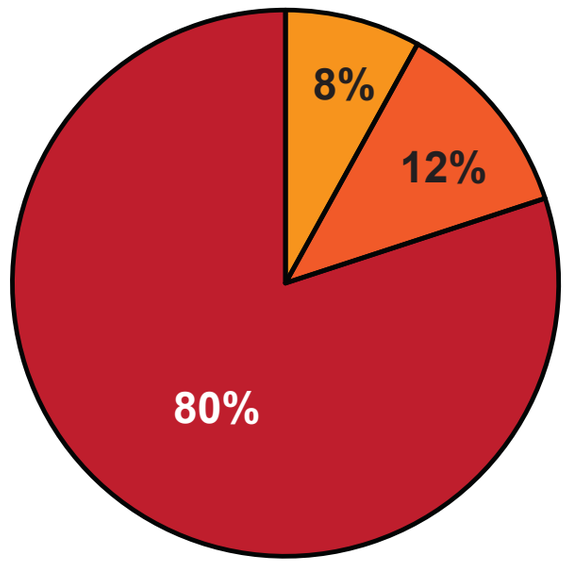
François Le Dily *et al*. Genes and Development (2014) in press

TADbit previous applications...

Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation

François Le Dily et al. Genes and Development (2014) in press

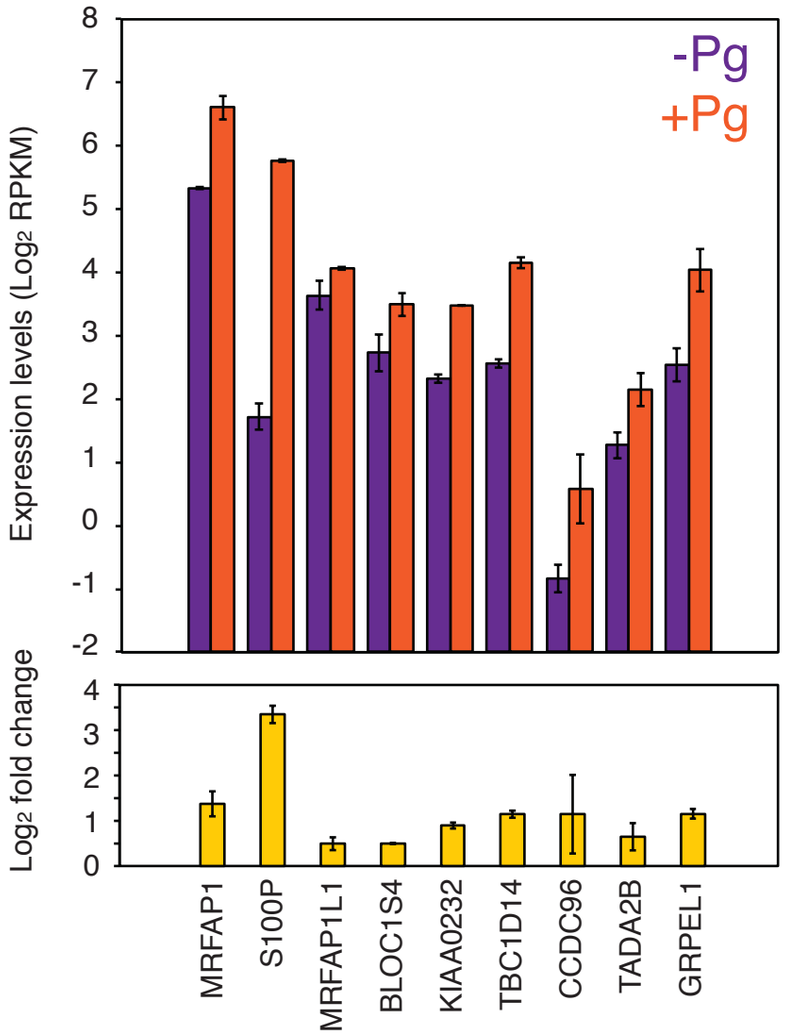


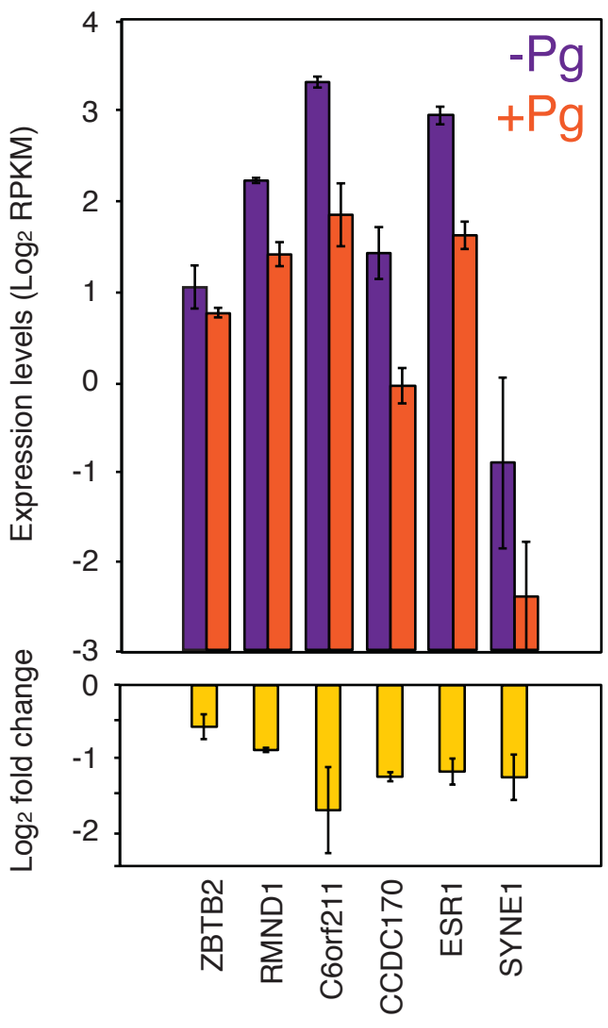


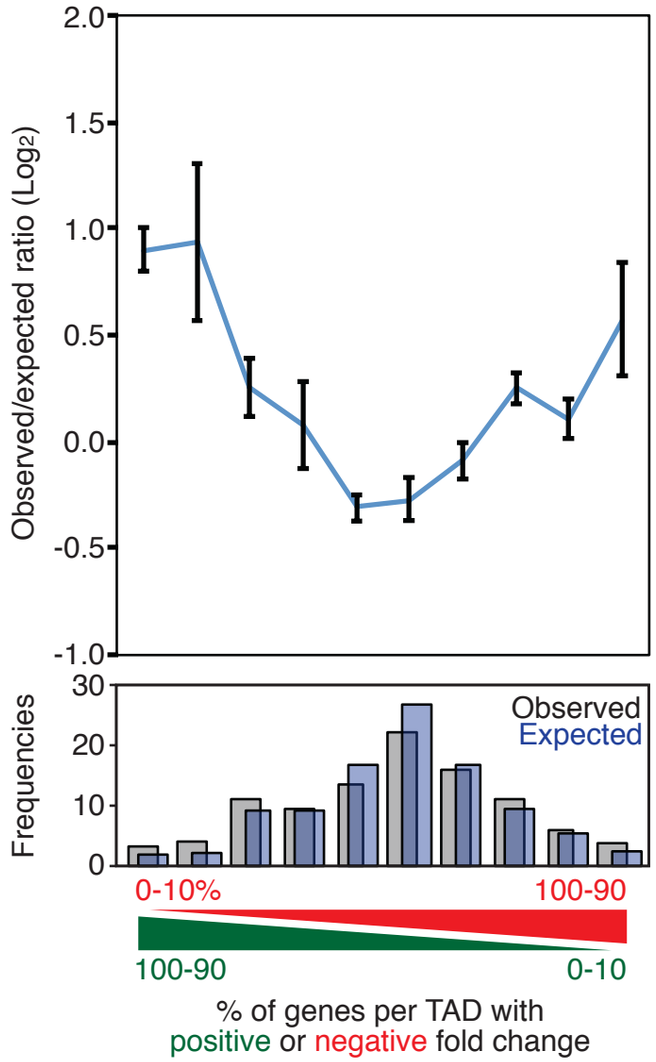
TADbit previous applications...

Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation

François Le Dily et al. Genes and Development (2014) in press





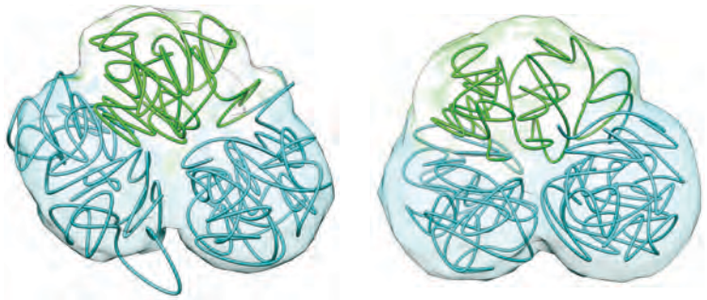


TADbit previous applications...

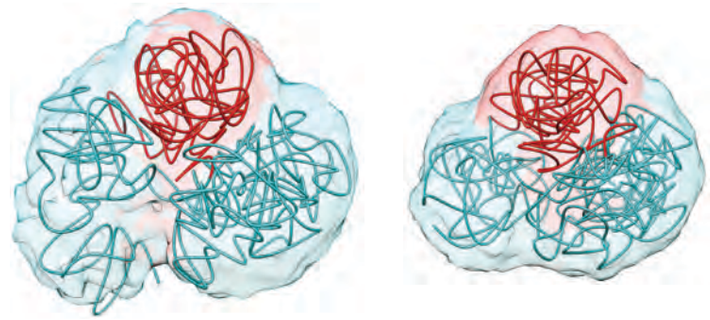
Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation

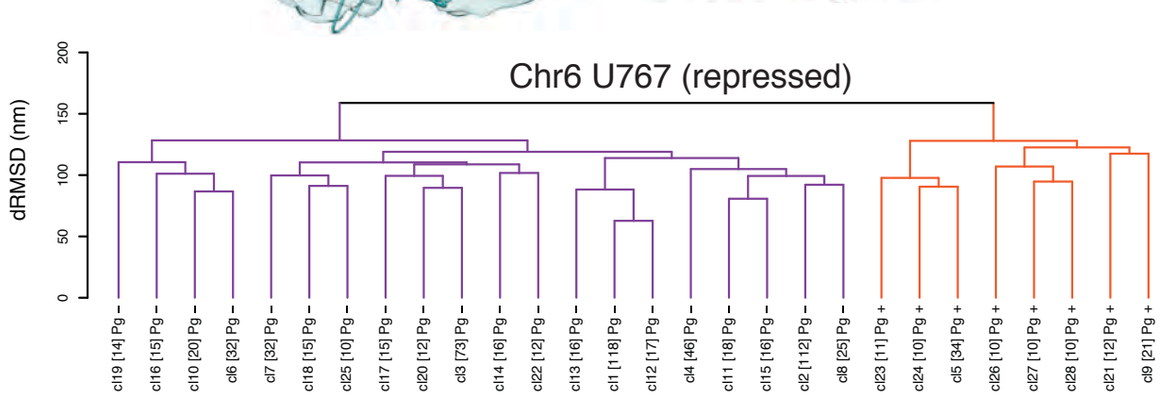
François Le Dily et al. Genes and Development (2014) in press

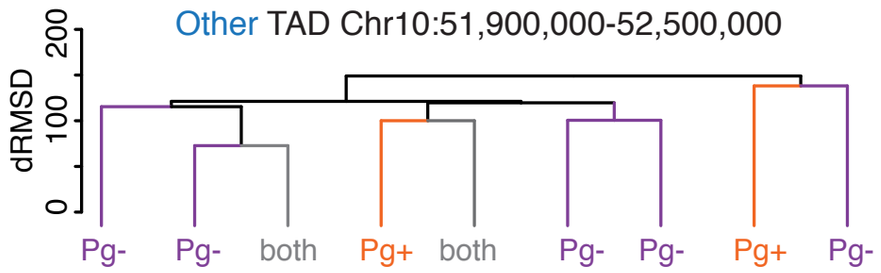
Chr2:9,600,000-13,200,000



Chr6:71,800,000-76,500,000



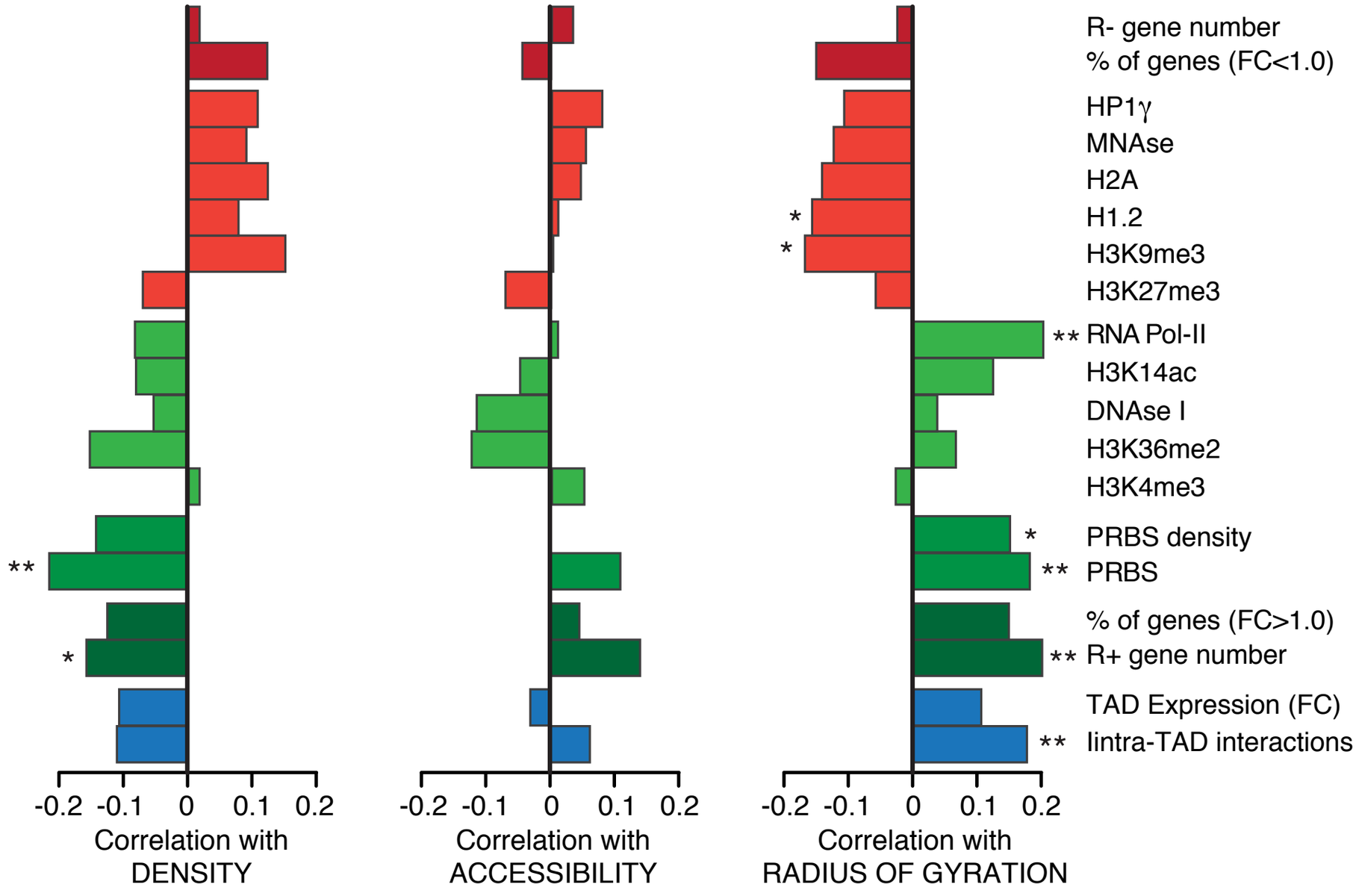




TADbit previous applications...

Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation

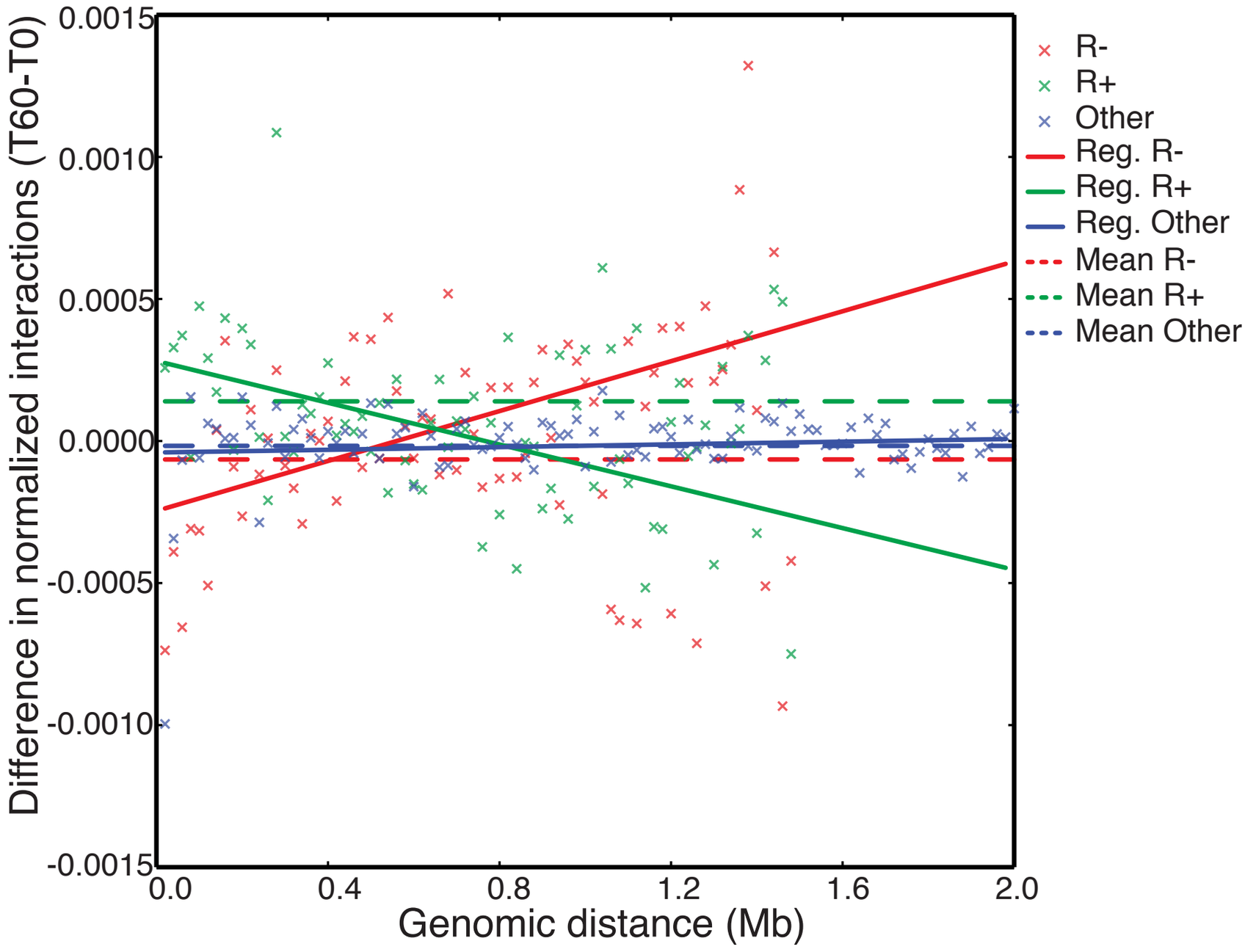
François Le Dily et al. Genes and Development (2014) in press



TADbit previous applications...

Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation

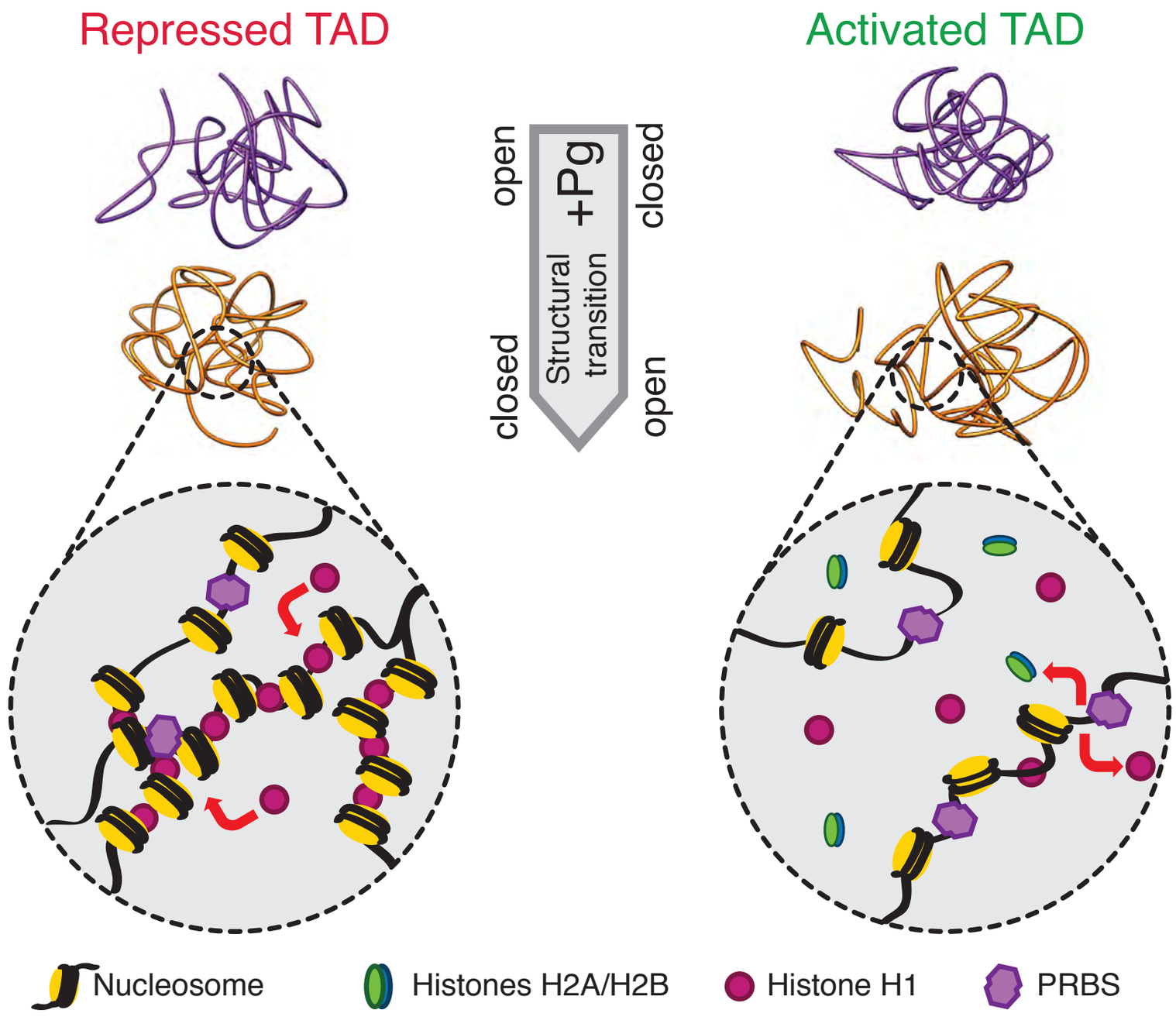
François Le Dily et al. Genes and Development (2014) in press



TADbit previous applications...

Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation

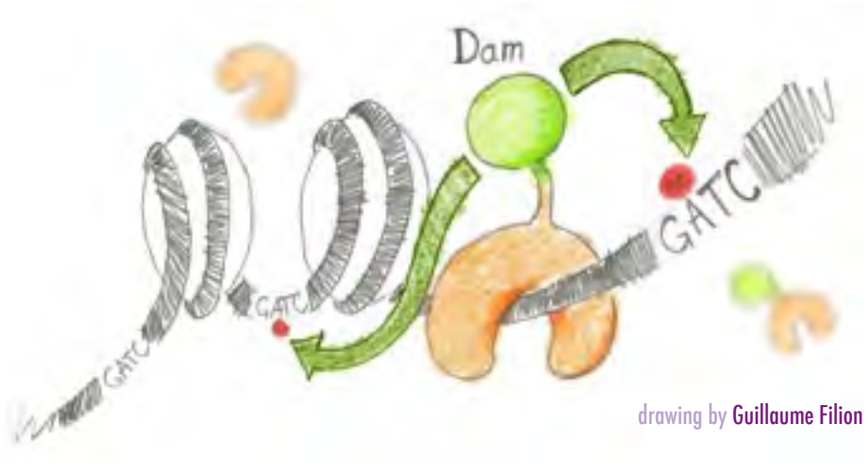
François Le Dily et al. Genes and Development (2014) in press



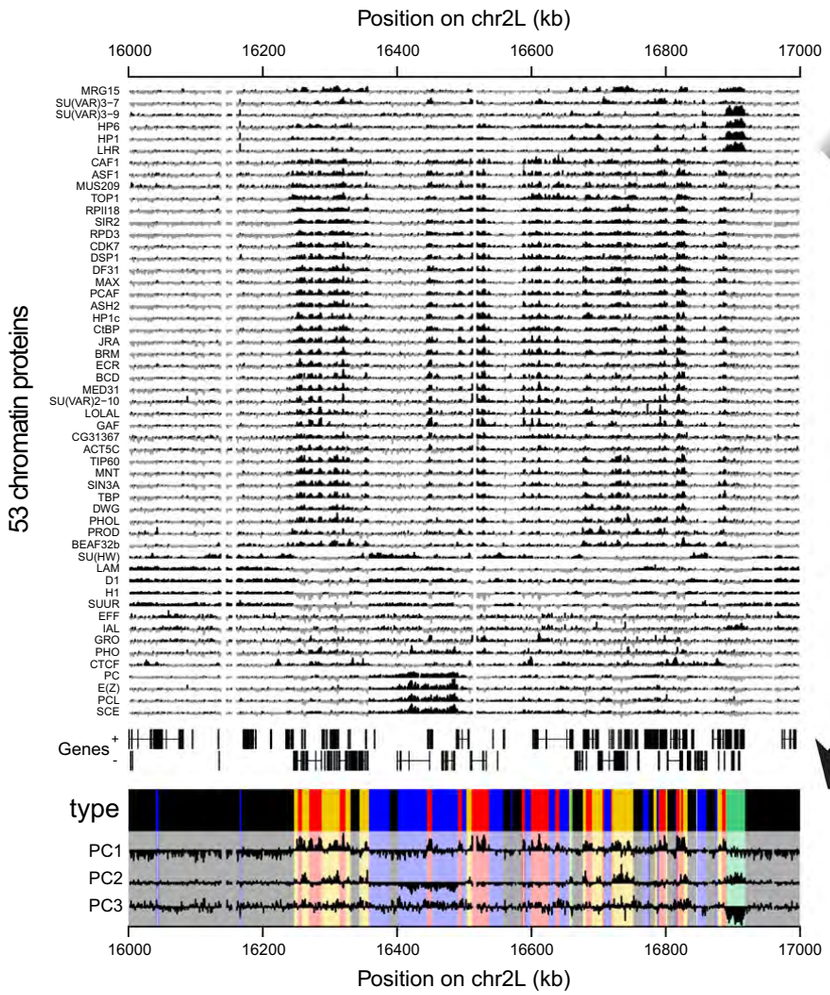
Structuring the COLORs of chromatin

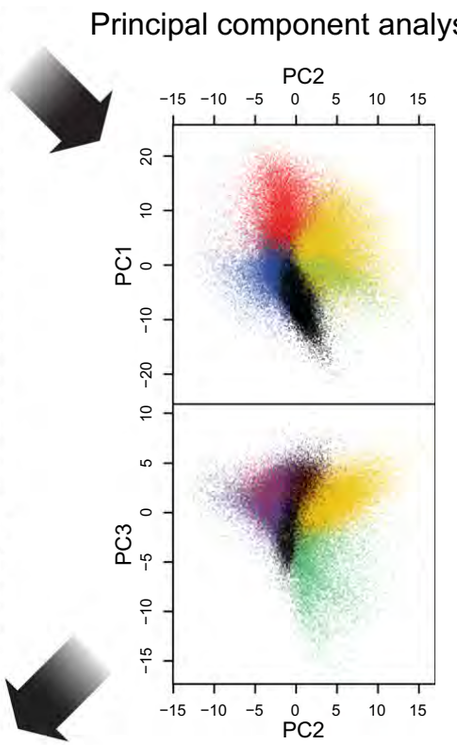
Fly Chromatin COLORs

Filion et al. (2010). Cell, 143(2), 212–224.

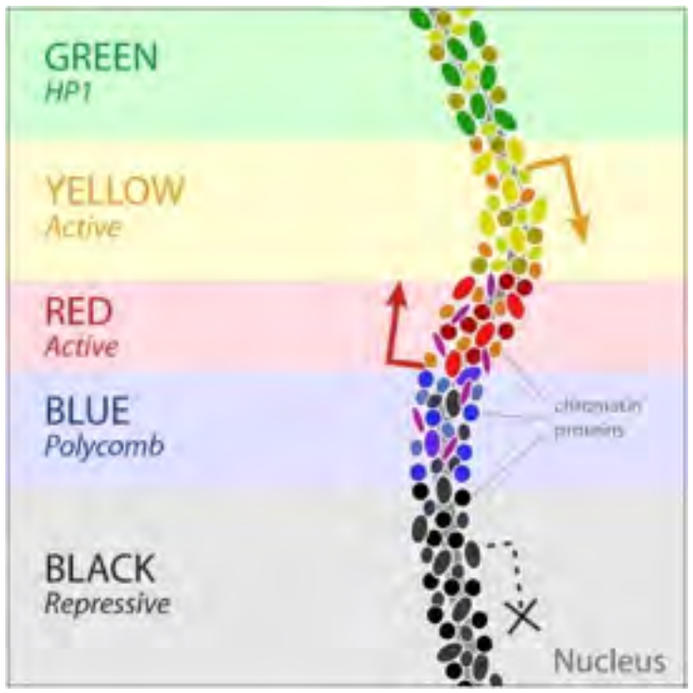


drawing by Guillaume Filion



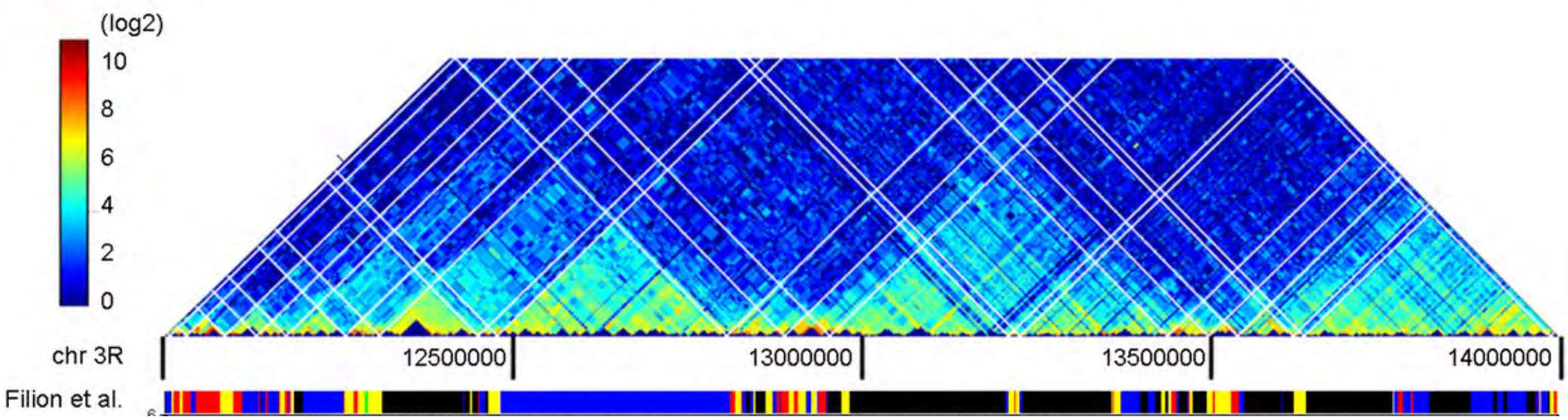


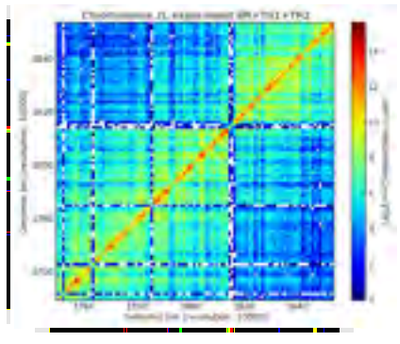
Hidden Markov model



Functional COLORs

Hou et al. (2012). Molecular Cell, 48(3), 471–484.





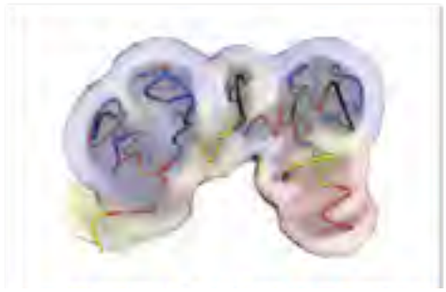
50 ~1Mb regions
10 for each color

Structural properties

50 1Mb regions. 10 enriched for each color.

RED dense region
3L:10610000-11630000

23.0% 17.8% 0% 42.8% 16.3%



YELLOW dense region
X:15590000-16600000

6.4% 43.5% 7.5% 18.9% 23.6%

GREEN dense region
2R:510000-1530000

0.1% 1.0% 98.3% 0% 0%

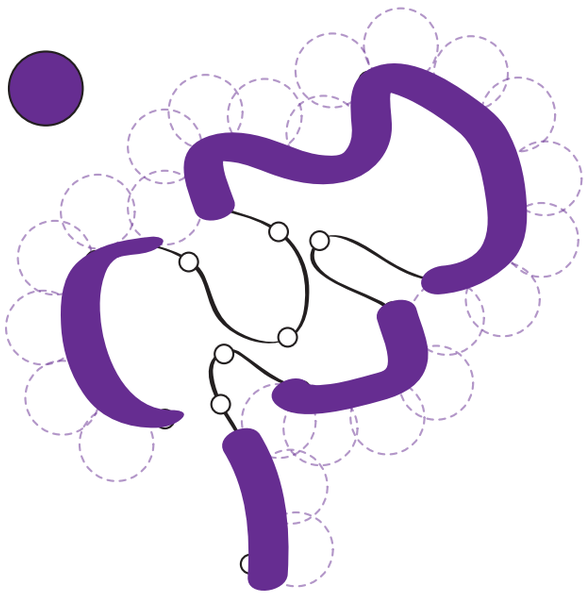
BLUE dense region
X:2340000-3370000

12.7% 11.3% 1.1% 50.6% 23.3%

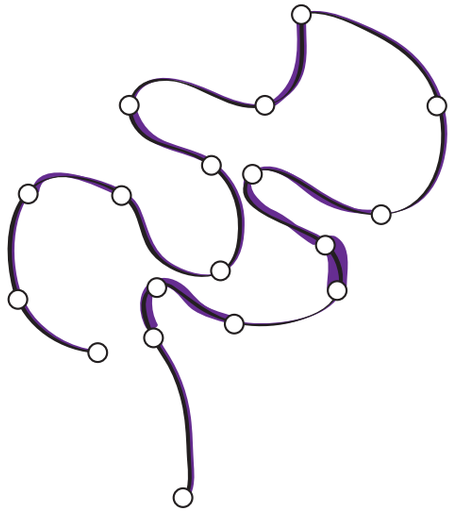
BLACK dense region
2L:17500000-18530000

2.0% 2.4% 1.6% 2.0% 92.0%

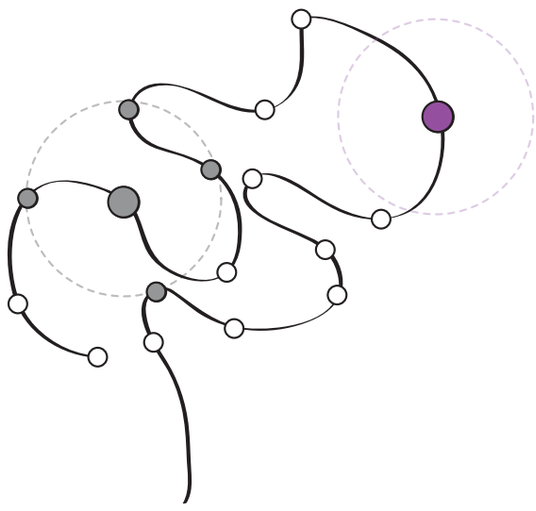
Accessibility (%)



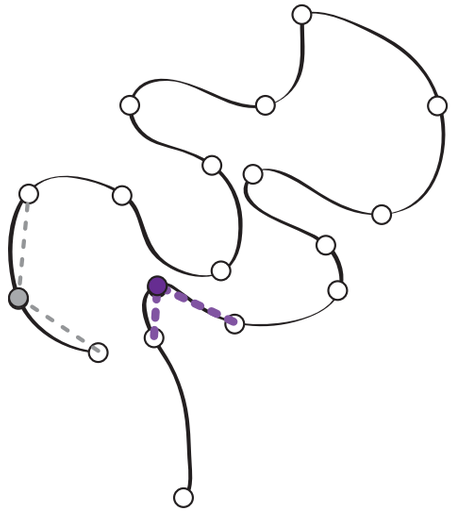
Density (bp/nm)



Interactions



Angle



Structural COLORs

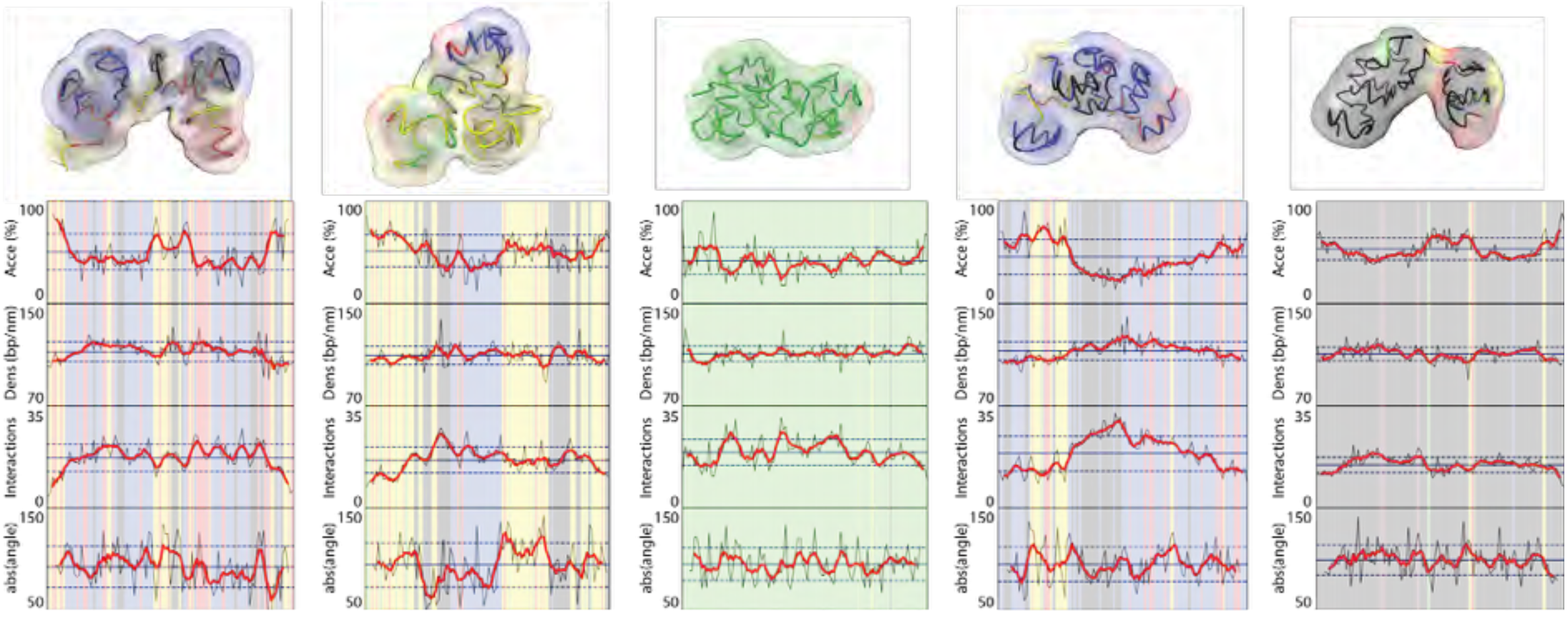
RED dense region
3L:10610000-11630000
23.0% 17.8% 0% 42.8% 16.3%

YELLOW dense region
X:15590000-16600000
6.4% 43.5% 7.5% 18.9% 23.6%

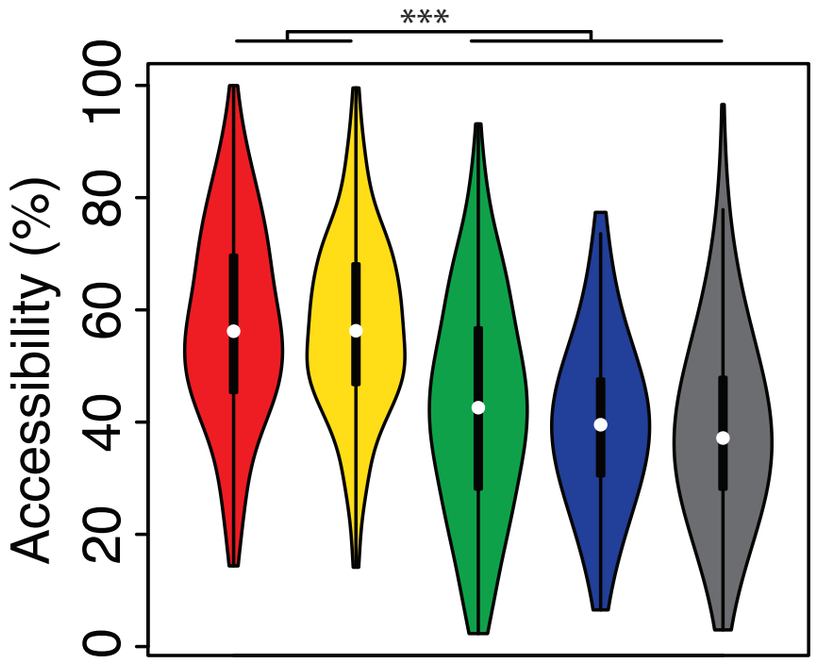
GREEN dense region
2R:510000-1530000
0.1% 1.0% 98.3% 0% 0%

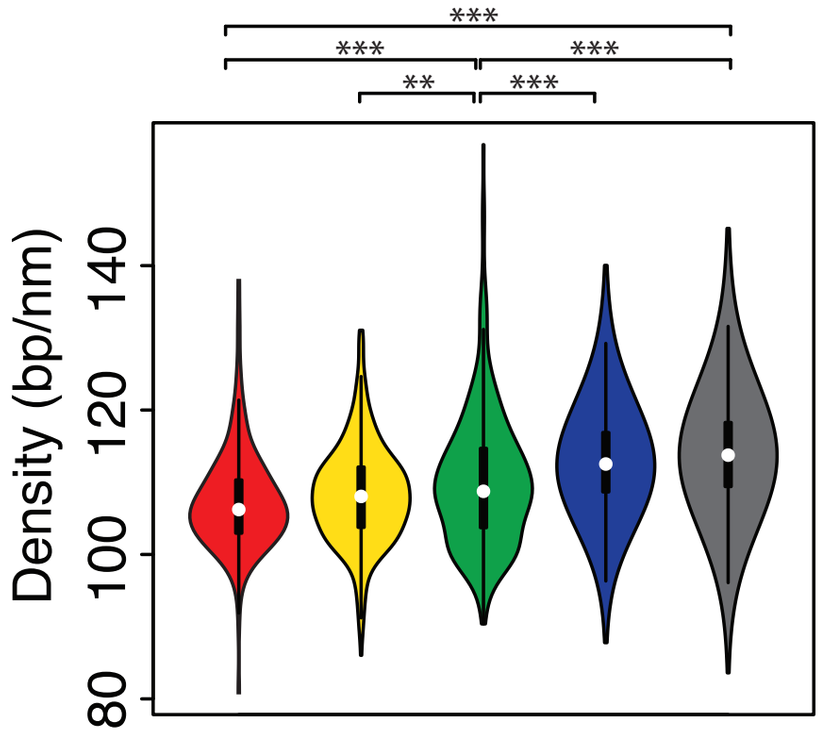
BLUE dense region
X:2340000-3370000
12.7% 11.3% 1.1% 50.6% 23.3%

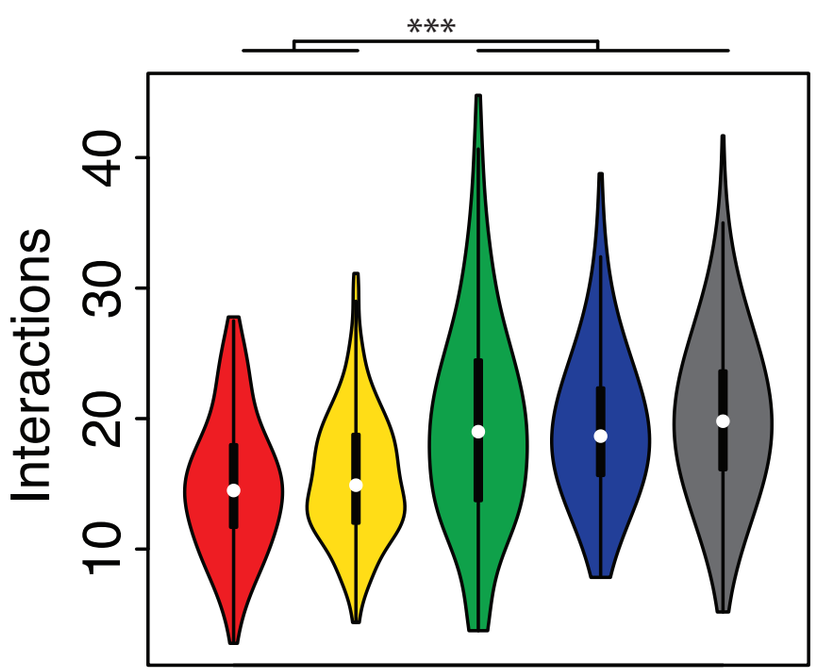
BLACK dense region
2L:17500000-18530000
2.0% 2.4% 1.6% 2.0% 92.0%

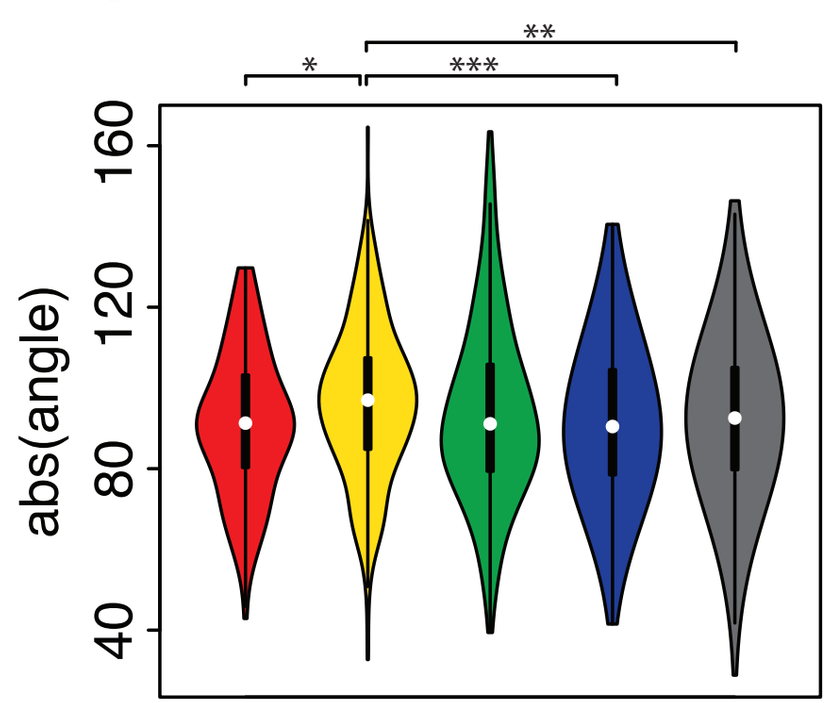


Structural COLORs

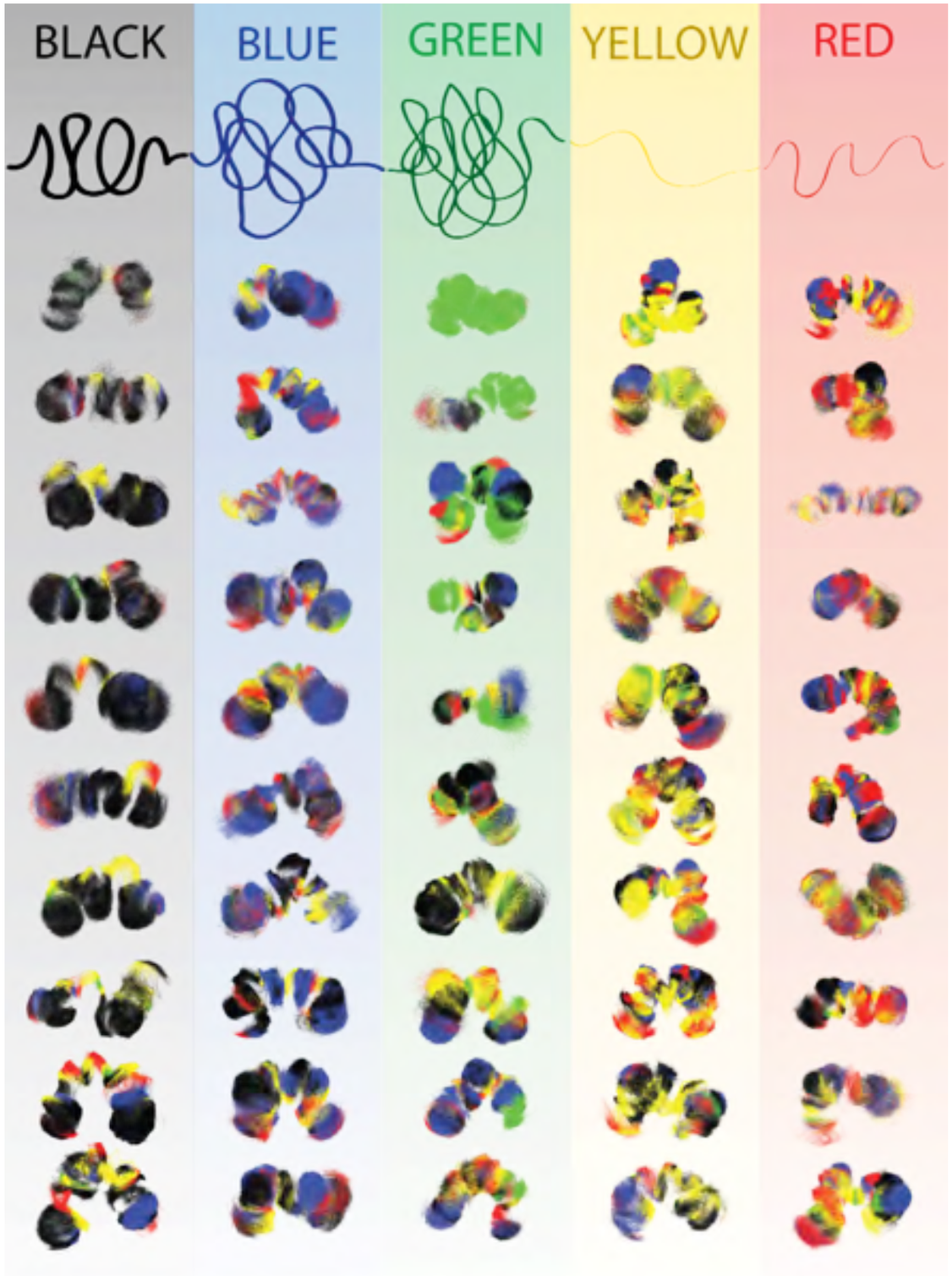


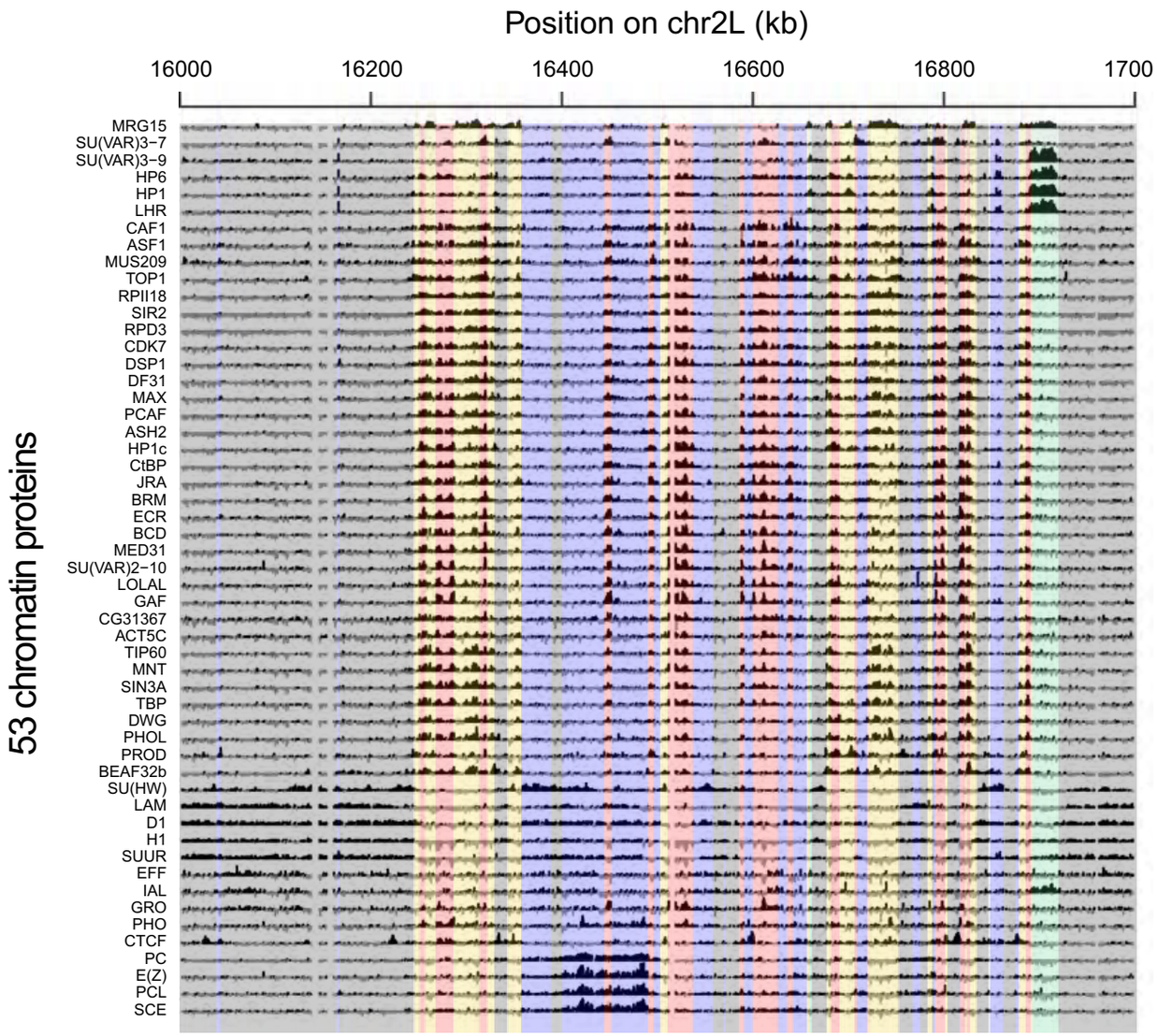






Structural COLORs

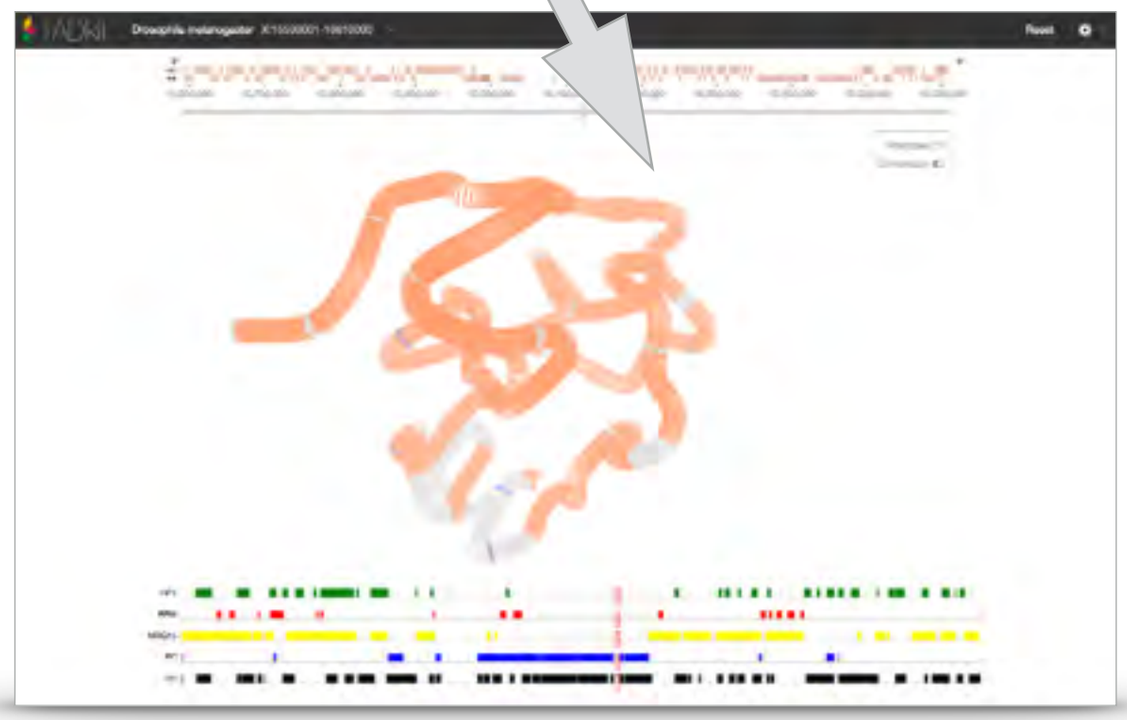




TADbit

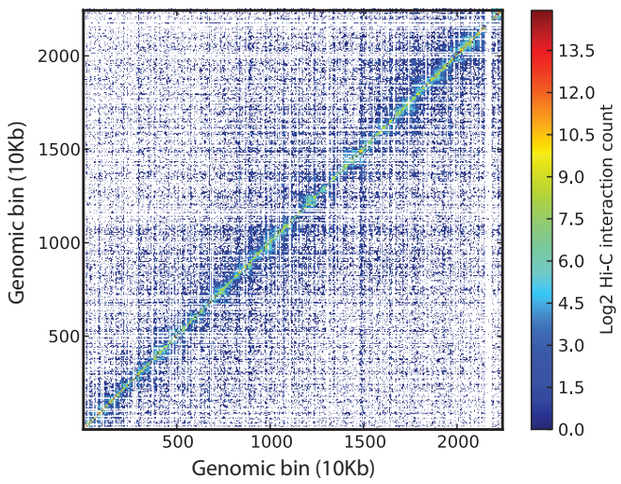
TADkit

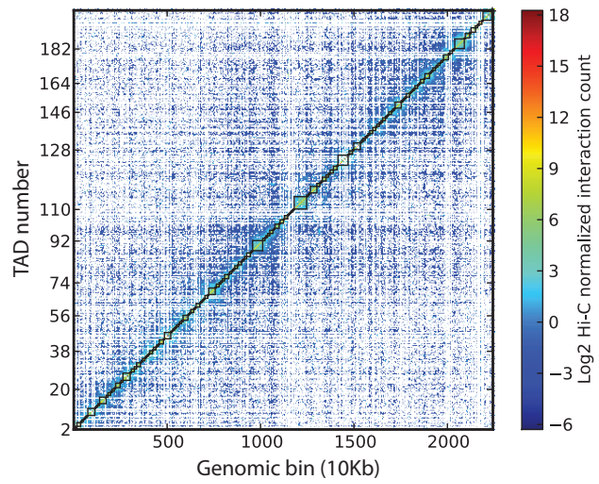
http://3DGenomes.org

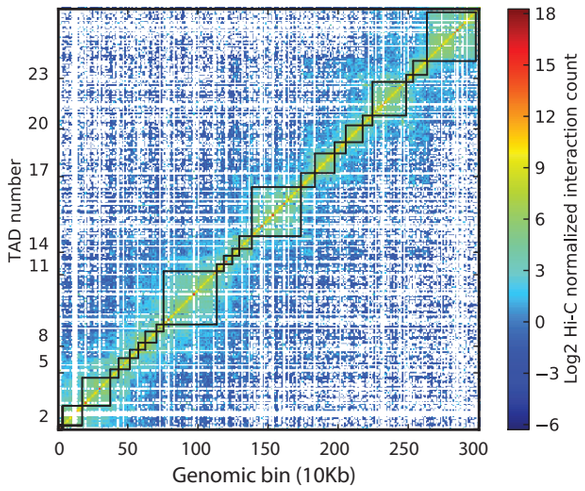


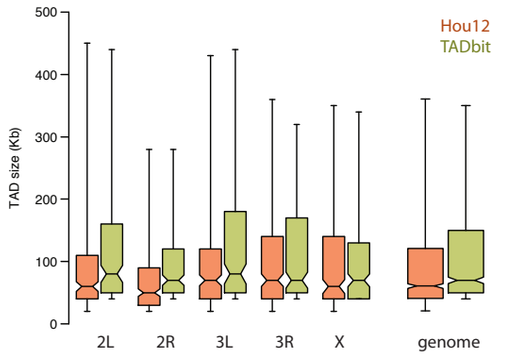
cnag CRG Centre for Genomic Regulation

TADbit matrix analysis





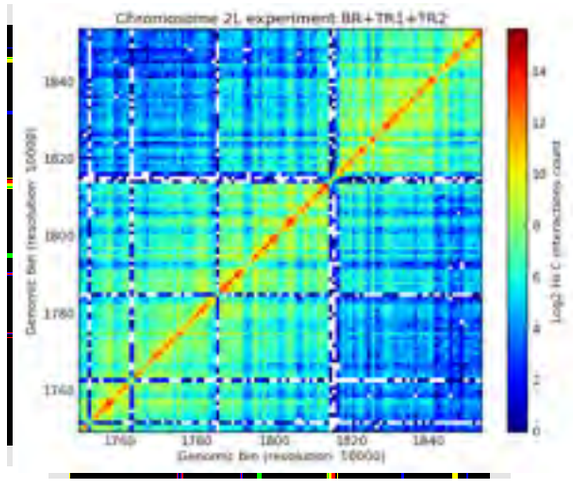


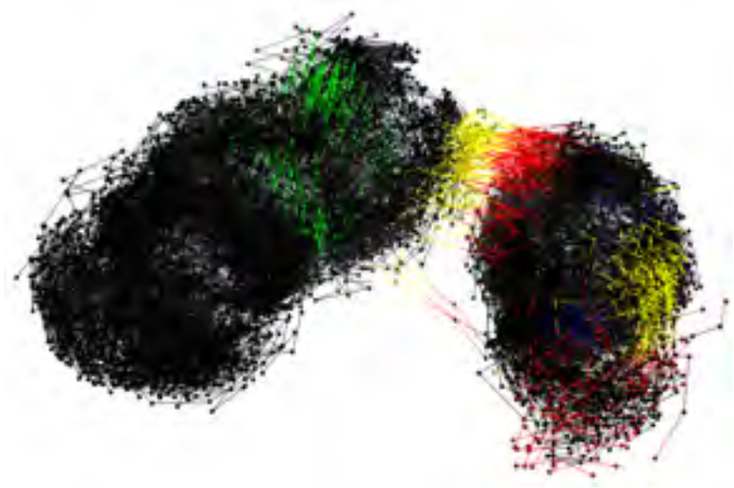



```python
#!/usr/bin/python
import sys
import os

# Libraries
from pytadbit import Chromosome  # Chromosome

# Create the chromosome object
my_chrom = Chromosome("X", species='Drosophila melanogaster', assembly='BDGP 5', )
# Read experimental data
my_chrom.add_experiment("MyExp", exp_type='Hi-C', enzyme='HindIII', # general descriptive fields
                        cell_type='kc167', identifier='GSE38468',   # general descriptive fields
                        article='TADBit',                           # user descriptions
                        resolution=100000,                          # Resolution of the input matrices
                        hic_data='my_matrix_10Kb.txt')              # File with the matrix

# Find TAD borders for each experiment
my_chrom.find_tad('MyExp', n_cpus=8)

# Visualize both experiments
my_chrom.visualize('MyExp', paint_tads=True, focus=(490,620), normalized=True)

# Align the TAD borders in both experiments and visualize
ali = my_chrom.alignment[('MyExp','Hou12')]
ali.draw()
```

cnag CRG Centre for Genomic Regulation

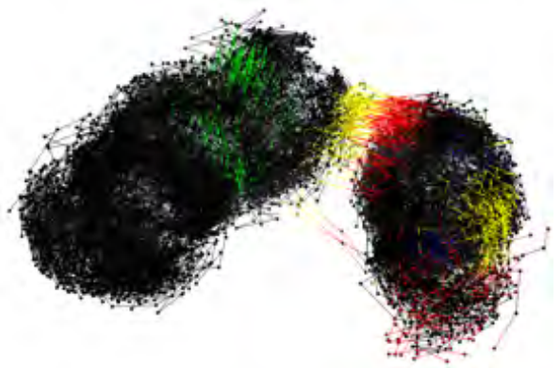
TADbit 3D modeling

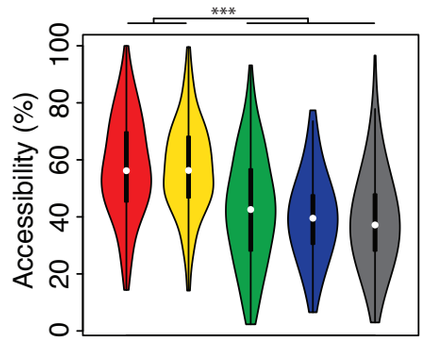


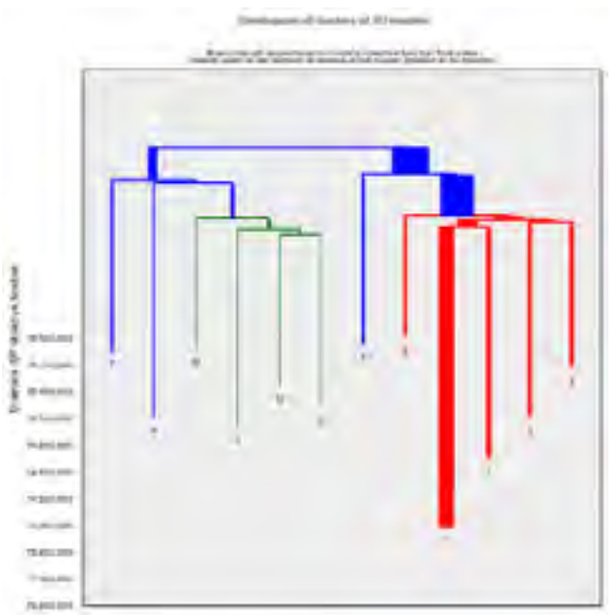


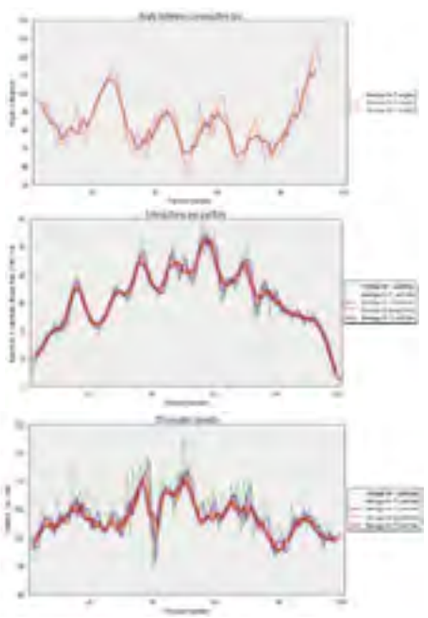
```python
# Create the chromosome object
my_chrom = Chromosome("X", species='Drosophila melanogaster', assembly='BDGP 5', )
# Read experimental data
my_chrom.add_experiment("MyExp", exp_type='Hi-C', enzyme='HindIII', # general descriptive fields
                        cell_type='kc167', identifier='GSE38468',   # general descriptive fields
                        article='TADBit',                           # user descriptions
                        resolution=100000,                          # Resolution of the input matrices
                        hic_data='my_matrix_10Kb.txt')              # File with the matrix

# Define optimal paramaters
optpar = {'kforce': 5,
          'lowfreq': -0.3,
          'lowrdist': 100,
          'upfreq': 0.6,
          'maxdist': 200,
          'scale': 0.01,
          'reference': 'Dm_BDGP5'}

# Build 3D models based on the HiC data.
models = exp.model_region(20000, 30000, n_models=10, n_keep=10, n_cpus=4, keep_all=True, config=optpar)

# Save your entire analysis and models
models.save_models('MyModels.models')
```

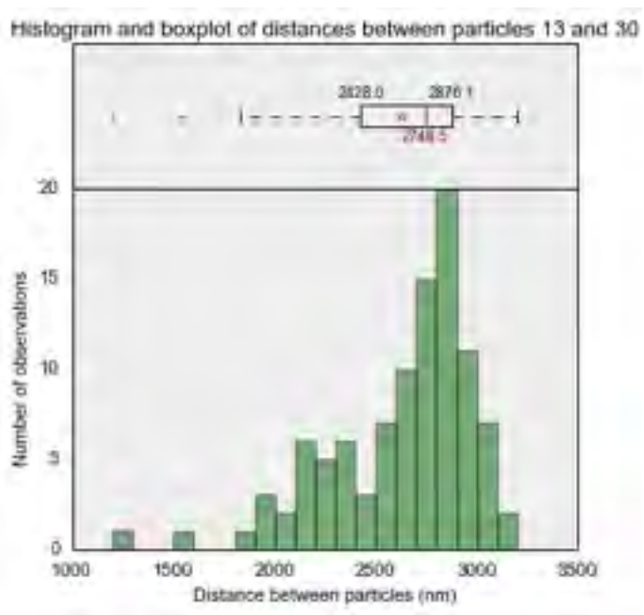
TADbit model analysis












```
# Clusters
models.cluster_analysis_dendrogram(color=True)

# Interactions, angle and density per particle
models.interactions(cutoff=2000)
models.walking_angle(steps=(3, 5, 7), signed=False)
models.density_plot()

# Plot the distance distributions between particles 13 and 30 in the top 100 models
models.median_3d_dist(13, 30, models=range(100))
```

TADkit demo


```
{
    "metadata": {
            "formatVersion" : 3,
            "article" : "Hnou et al. (2012) CELL",
            "species" : "Drosophila melanogaster",
            "assembly" : "BDGP 5",
            "cellType" : "kc167",
            "experimentType" : "Hi-C",
            "identifier" : "GSE38468",
            "resolution" : 10000,
            "chromosome" : "X",
            "start" : 15590001,
            "end" : 16610000
        },

    "materials": [  {
            "DbgColor" : 15658734,
            "DbgIndex" : 0,
            "DbgName" : "default",
            "vertexColors" : false,
            "colorAmbient" : [0.0, 0.0, 0.0],
            "colorDiffuse" : [0.8, 0.8, 0.8],
            "colorSpecular" : [0.8, 0.8, 0.8],
            "illumination" : 2,
            "specularCoef" : 0.0,
            "transparency" : 1.0
        }],

    "vertices": [
            -4997.512, -14445.484, 6862.561,
            -4930.434, -14470.804, 6934.82,
            -4970.555, -14415.34, 7002.625,
            -4935.936, -14318.296, 7005.901,
            -4976.995, -14358.728, 7075.177,
            -4884.587, -14327.883, 7075.655,
            -4930.424, -14392.395, 7137.914,
            -4850.193, -14408.212, 7102.532,
            -4856.203, -14347.263, 7167.829,
            -4776.328, -14380.24, 7156.217,
            -4821.549, -14452.505, 7203.907,
            -4852.323, -14541.284, 7187.383,
            -4907.888, -14583.206, 7119.326,
            -4919.553, -14541.712, 7027.894,
            -4848.228, -14596.95, 7055.235,
            -4769.283, -14539.379, 7055.179,
            -4788.784, -14470.629, 7082.837,
            -4698.692, -14520.434, 7010.915,
            -4750.675, -14508.357, 6928.896,
            -4734.381, -14436.217, 6980.335,
            -4757.711, -14440.468, 6904.083,
            -4748.483, -14357.549, 6950.223,
            -4784.942, -14286.947, 6980.46,
            -4810.042, -14241.789, 6897.86,
            -4751.912, -14241.125, 6966.872,
            -4829.174, -14317.494, 6896.449,
```


TADkit


```
{
    "metadata": {
            "formatVersion" : 3,
            "article" : "Hnou et al. (2012) CELL",
            "species" : "Drosophila melanogaster",
            "assembly" : "BDGP 5",
            "cellType" : "kc167",
            "experimentType" : "Hi-C",
            "identifier" : "GSE38468",
            "resolution" : 10000
```


TADbit

Nov 24th-28th Lisbon

FUNDAÇÃO CALOUSTE GULBENKIAN

Instituto Gulbenkian de Ciência

http://gtpb.igc.gulbenkian.pt

cnag CRG

TADbit

Davide Baù

François Serra

TADkit

Mike Goodstadt

Yasmina Cuartero

Yannick Spill

Marco di Stefano

Acknowledgments

François Serra
Davide Baù
Mike Goodstadt

François le Dily
Yasmina Cuartero
Francisco Martínez-Jiménez
David Dufour
Gireesh Bogu

Job Dekker
Program in Systems Biology
Department of Biochemistry and Molecular Pharmacology
University of Massachusetts Medical School
Worcester, MA, USA

Kerstin Bystricky
Chromatin and gene expression
Laboratoire de Biologie Moléculaire Eucaryote - CNRS
Toulouse, France

Miguel Beato & Guillaume Filion
Gene Regulation, Stem Cells and Cancer
Centre de Regulació Genòmica
Barcelona, Spain

http://marciuslab.org
http://3DGenomes.org
http://cnag.cat · http://crg.cat