

# ***Boost-HiC : Computational enhancement of long-range contacts in chromosomal contact maps***



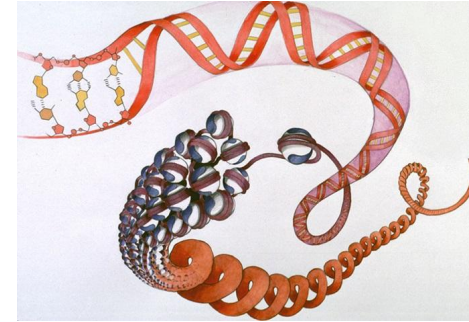
# The functional goal of 3D chromatin folding

## Sequence

...ACTGCCGCGTTGTCA...



## 3D Structure

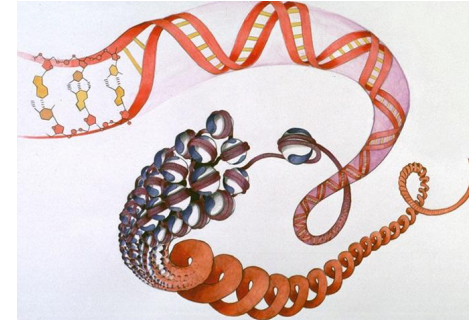


# The functional goal of 3D chromatin folding

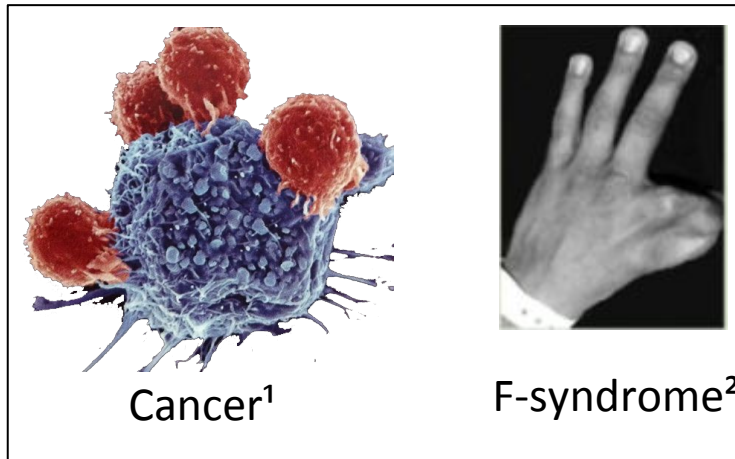
Sequence

3D Structure

...ACTGCCGCGTTGTCA...



Genetic disease



Cancer<sup>1</sup>

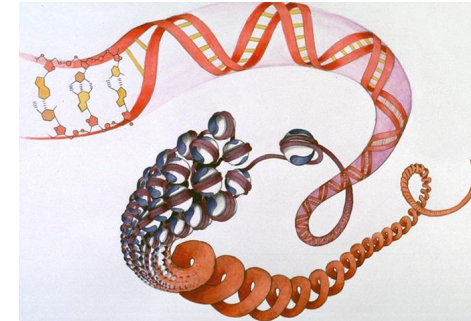
F-syndrome<sup>2</sup>

# The functional goal of 3D chromatin folding

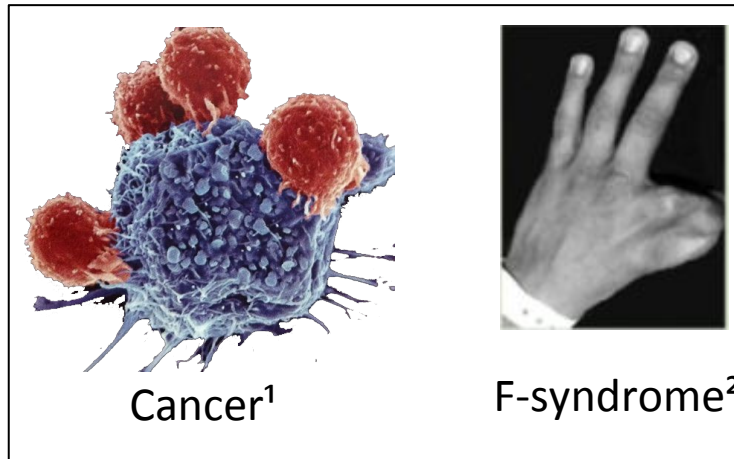
Sequence

3D Structure

...ACTGCCGCGTTGTCA...



Genetic disease



Study the 3D genomic structure at the best resolution

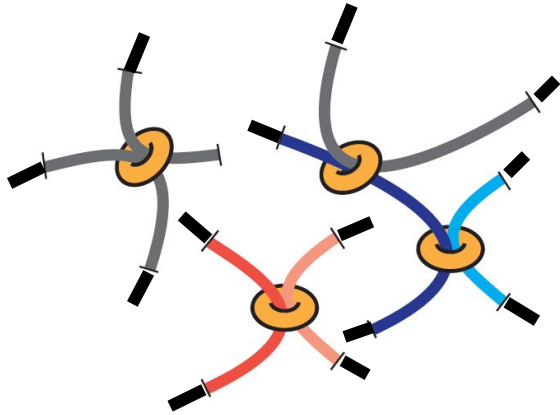
1 : Taberlay & A.I Genome Research.

2 : Lupianez & Al. Cell

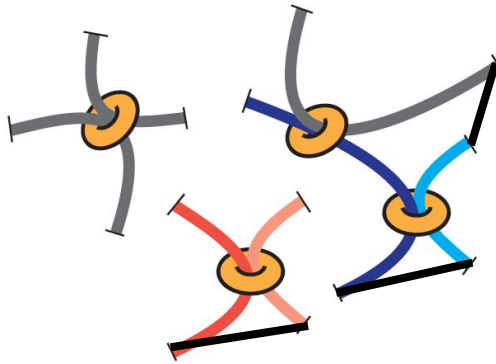
# High chromosome conformation capture (Hi-C)

## Protocol

Cross-link and digestion



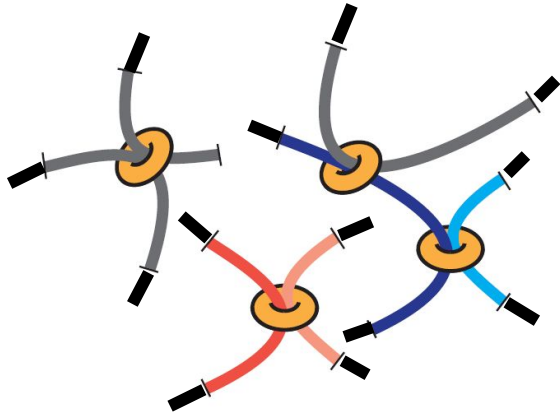
Ligation



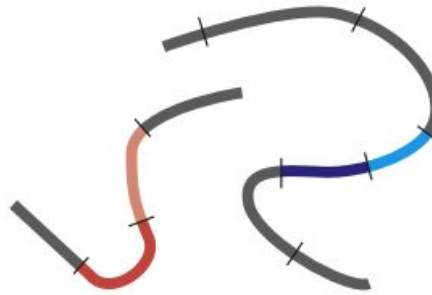
# High chromosome conformation capture (Hi-C)

## Protocol

Cross-link and digestion

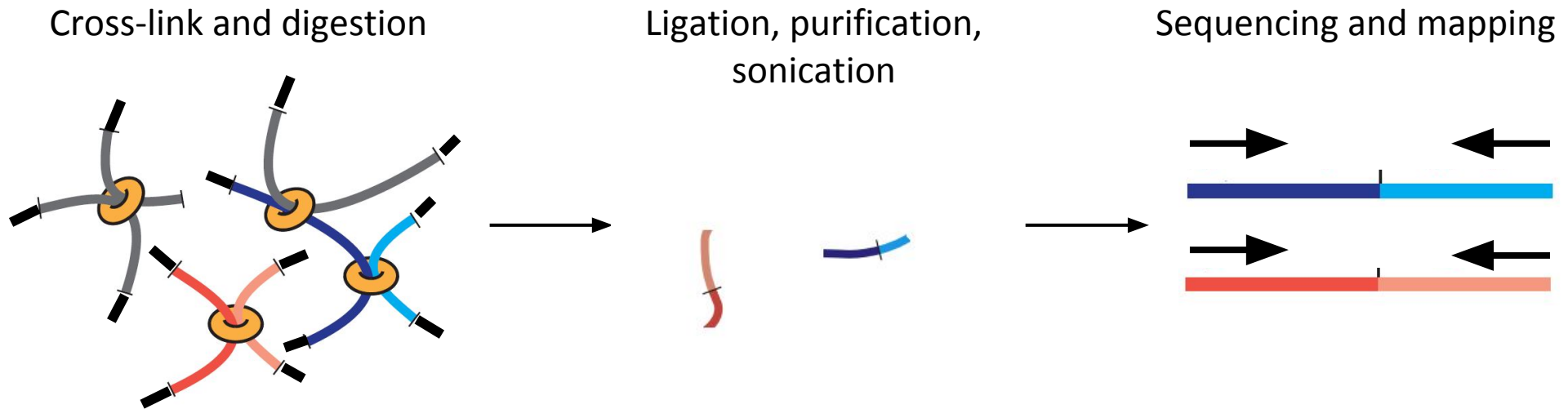


Ligation, purification



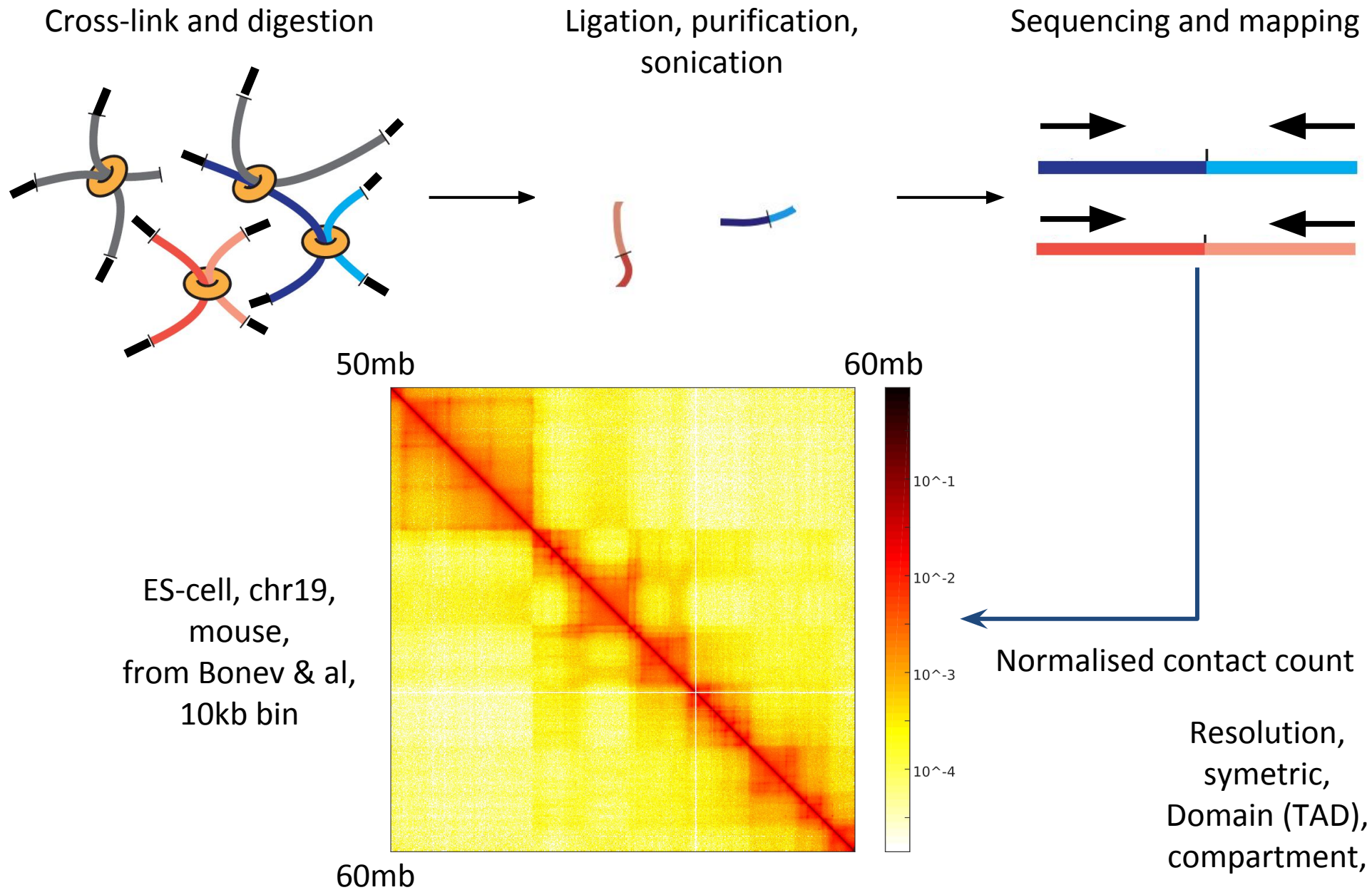
# High chromosome conformation capture (Hi-C)

## Protocol



# High chromosome conformation capture (Hi-C)

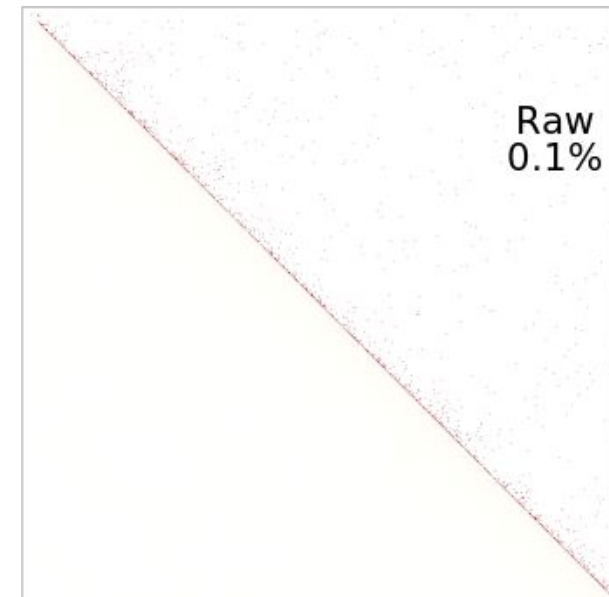
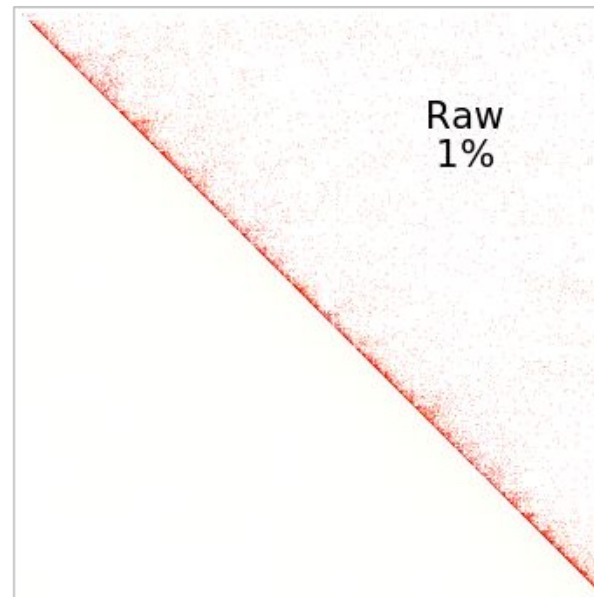
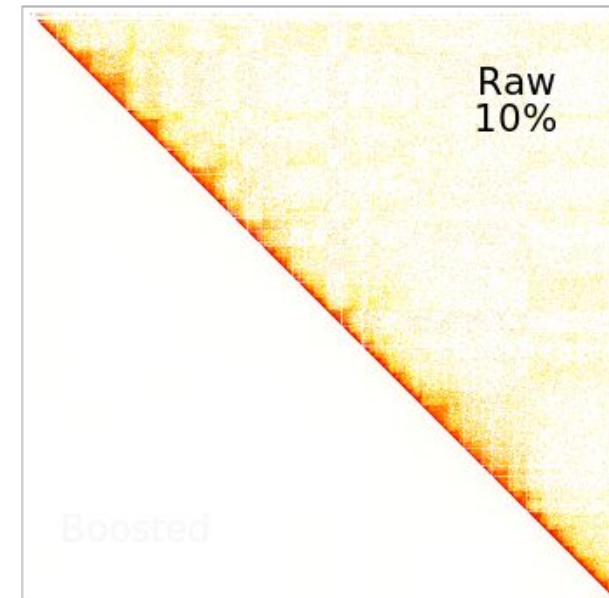
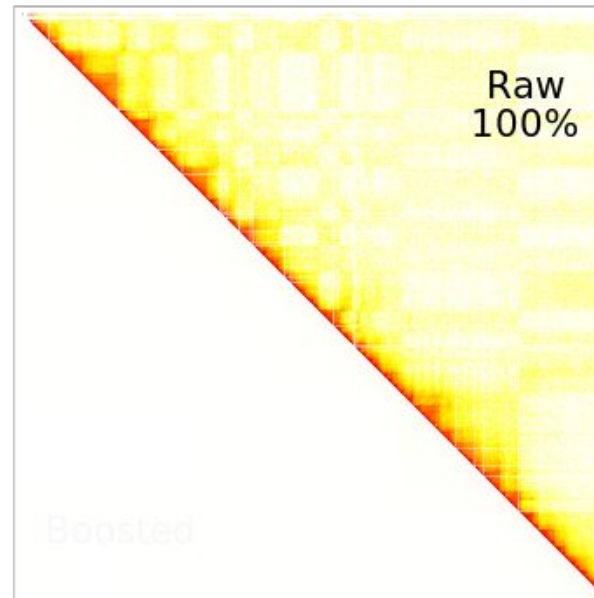
## Protocol



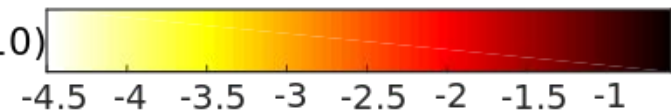


# Enhance the signal

chr16, ES-cell,  
100kb-29,8Mb

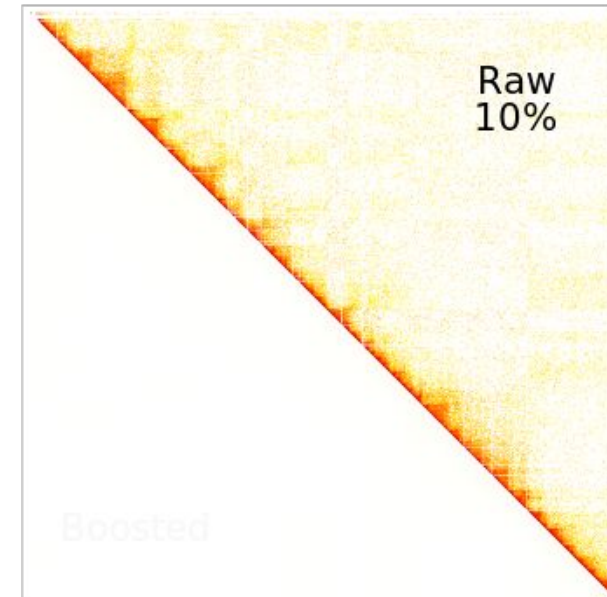
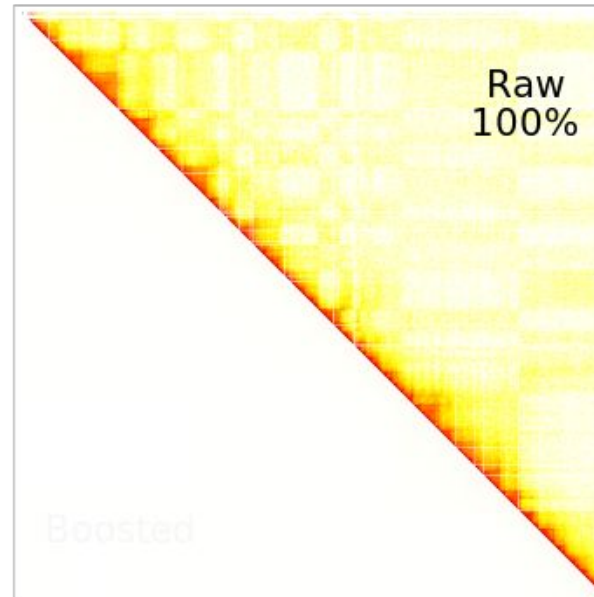


Contact frequency (log10)

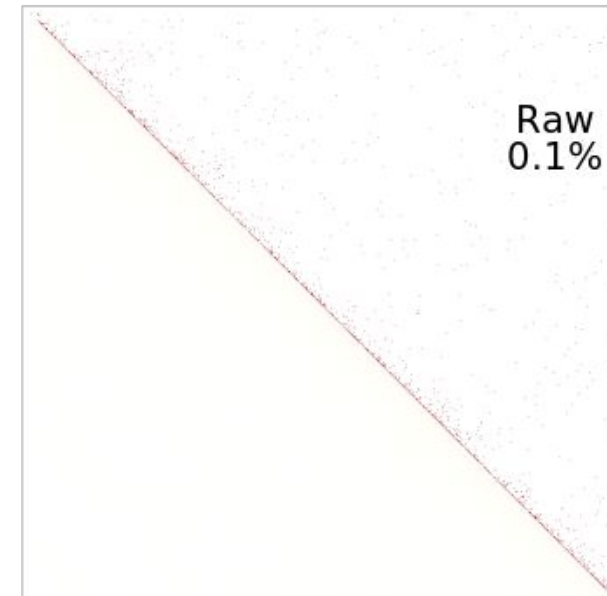
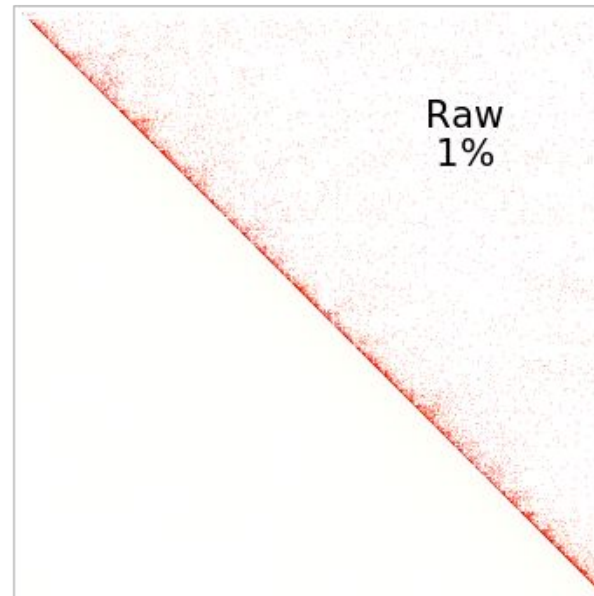


## Enhance the signal

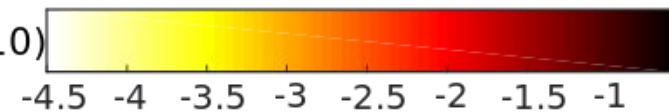
chr16, ES-cell,  
100kb-29,8Mb



Can we impute the missing  
signal from the contact map?



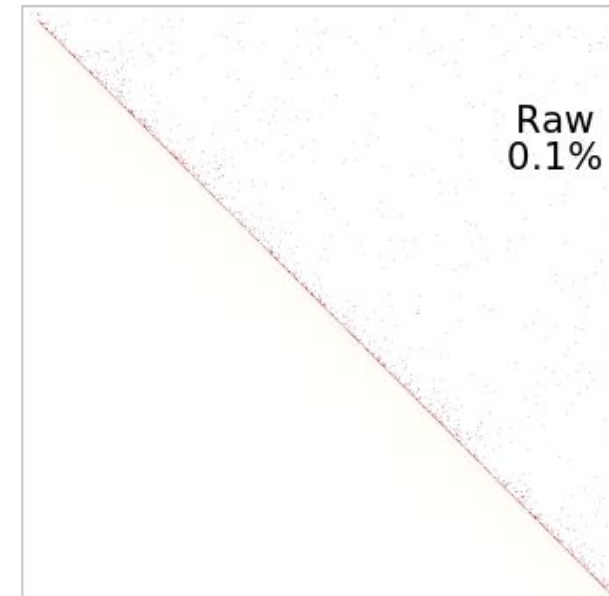
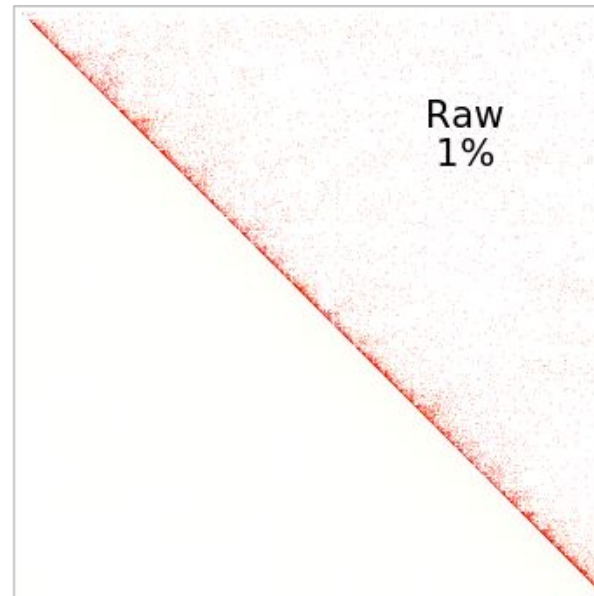
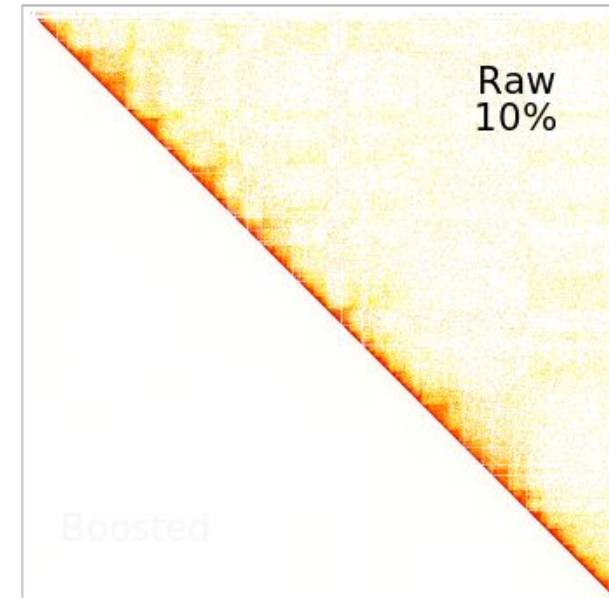
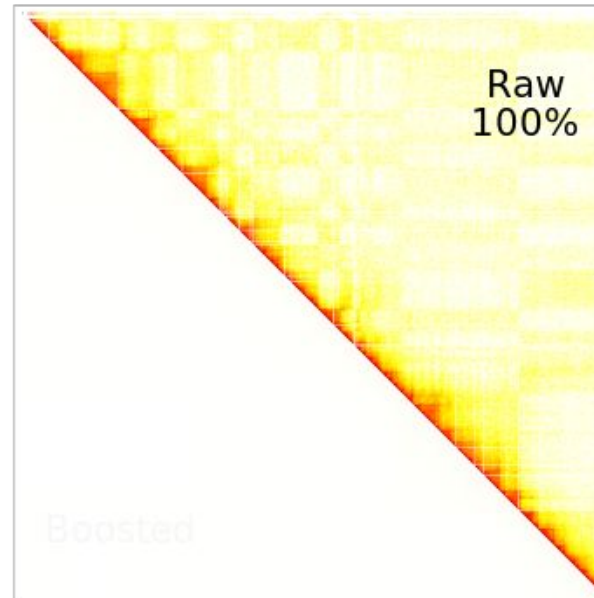
Contact frequency (log10)



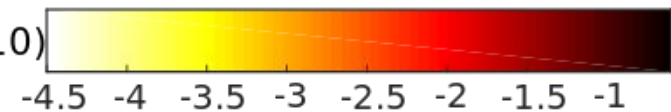
# Enhance the signal

chr16, ES-cell,  
100kb-29,8Mb

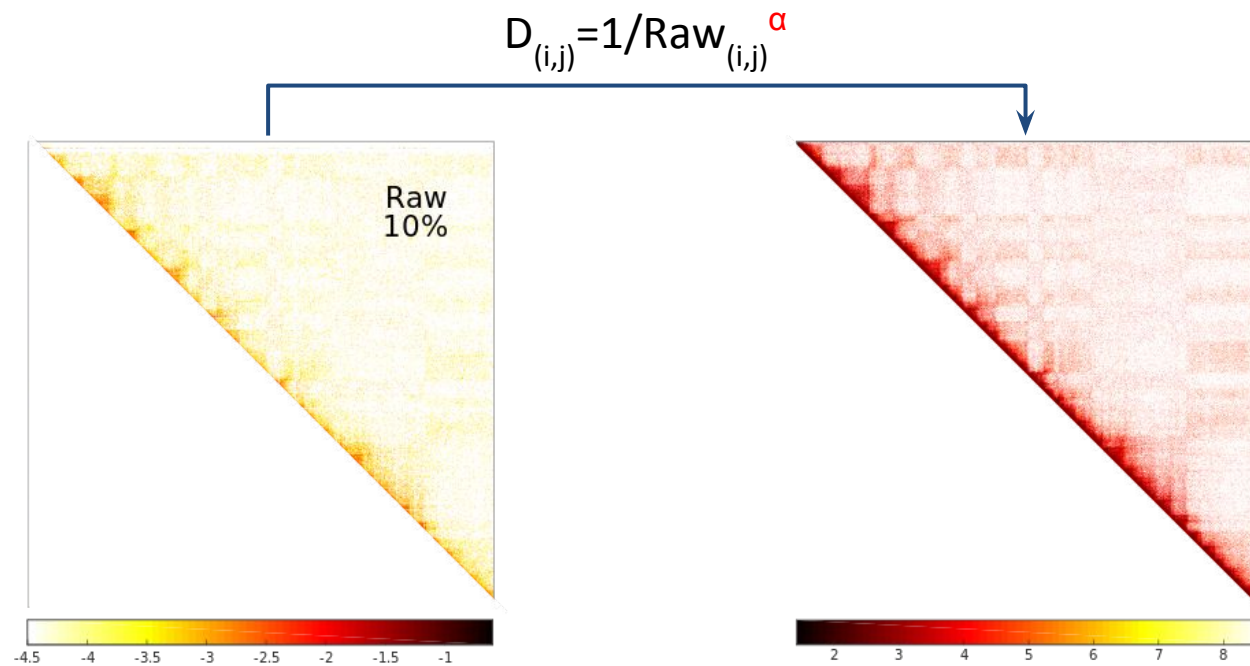
Can we impute the missing  
signal from the contact map?  
=> Boost-HiC



Contact frequency (log10)

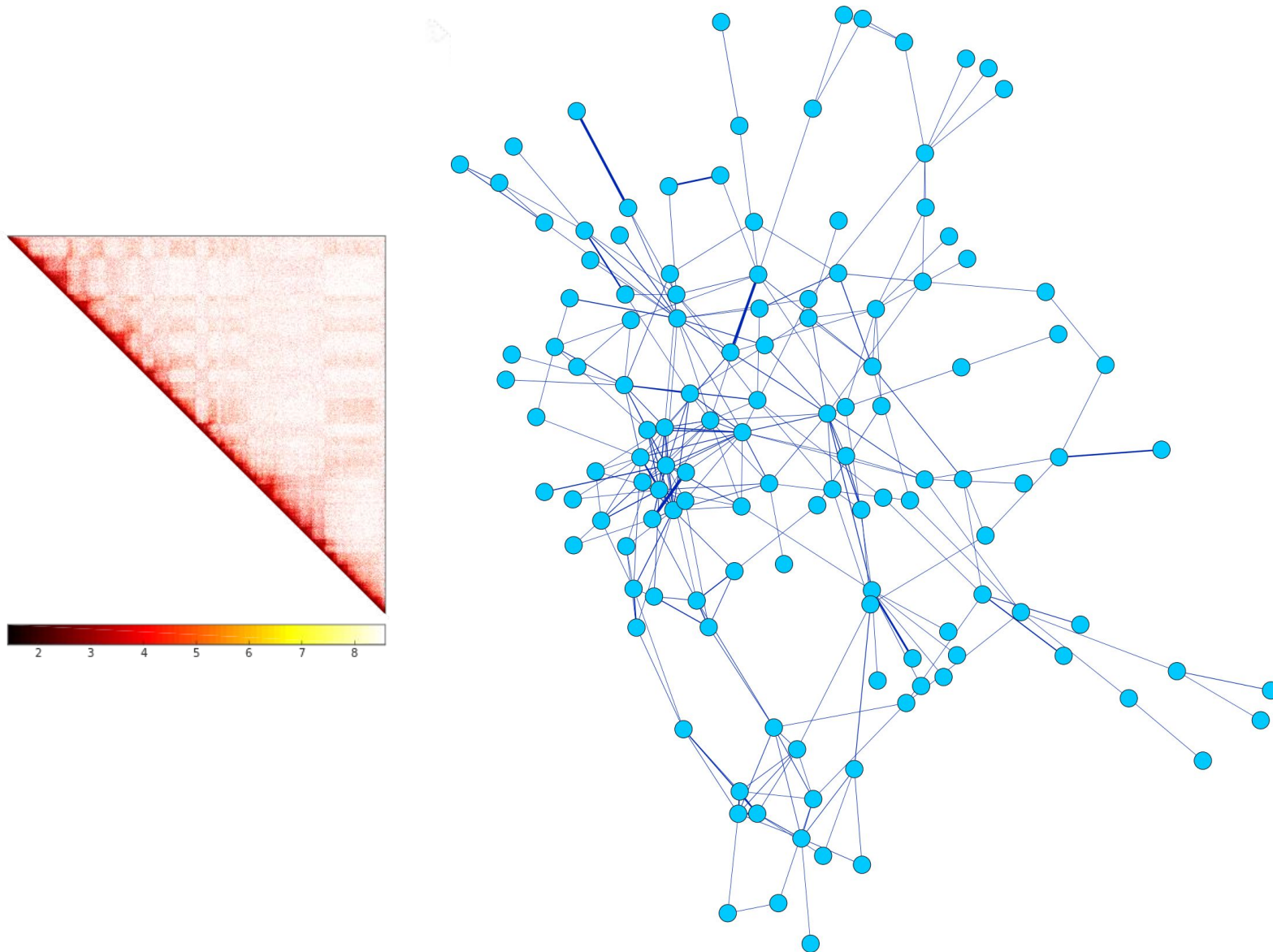


# Boost-HiC general procedure



transformation element wise  
color scale inversion  
Graph theory

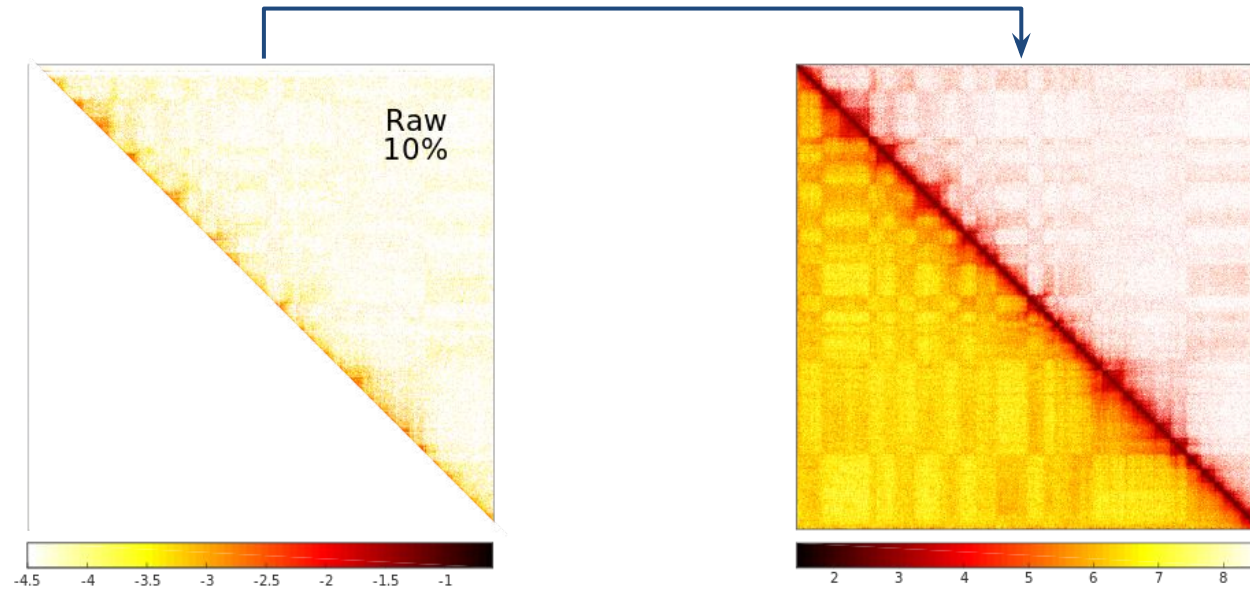
# Boost-HiC general procedure



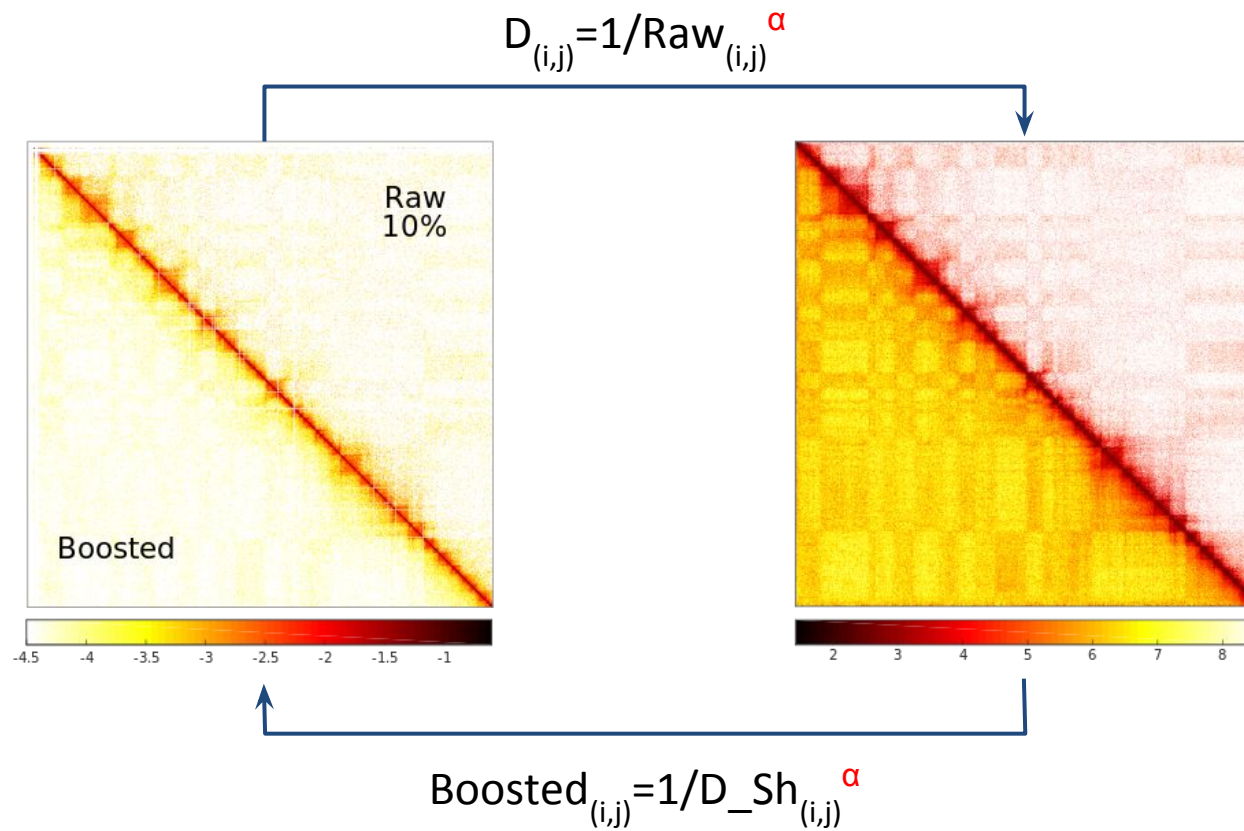


# Boost-HiC general procedure

$$D_{(i,j)} = 1/\text{Raw}_{(i,j)}^{\alpha}$$

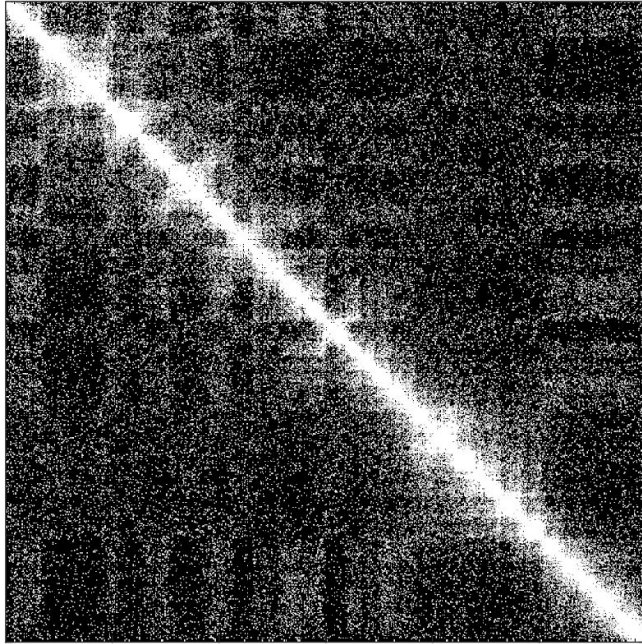


# Boost-HiC general procedure

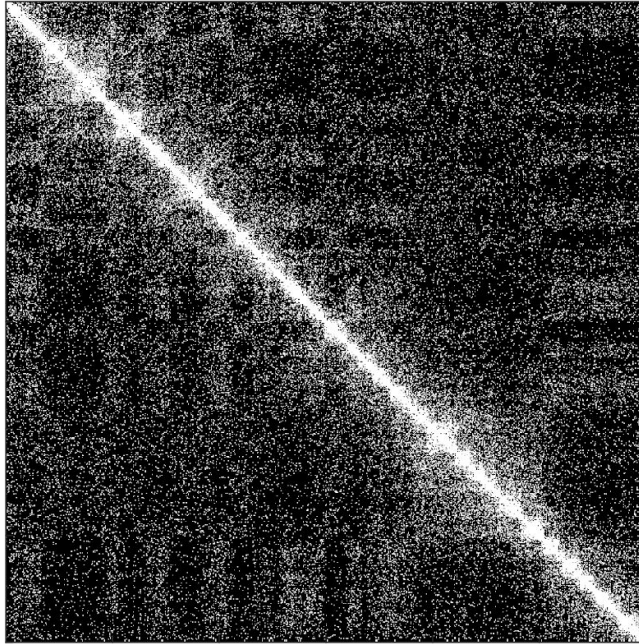


# Optimising the alpha parameter

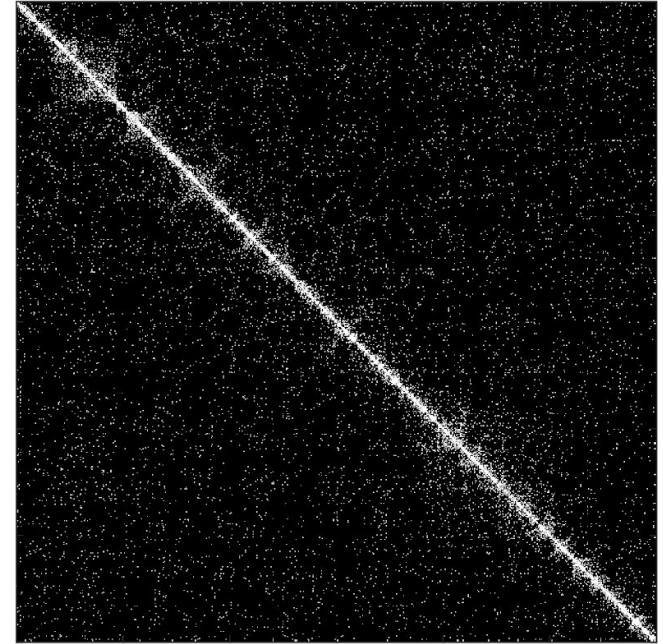
$\alpha = 0.18$



$\alpha = 0.32$



$\alpha = 0.6$



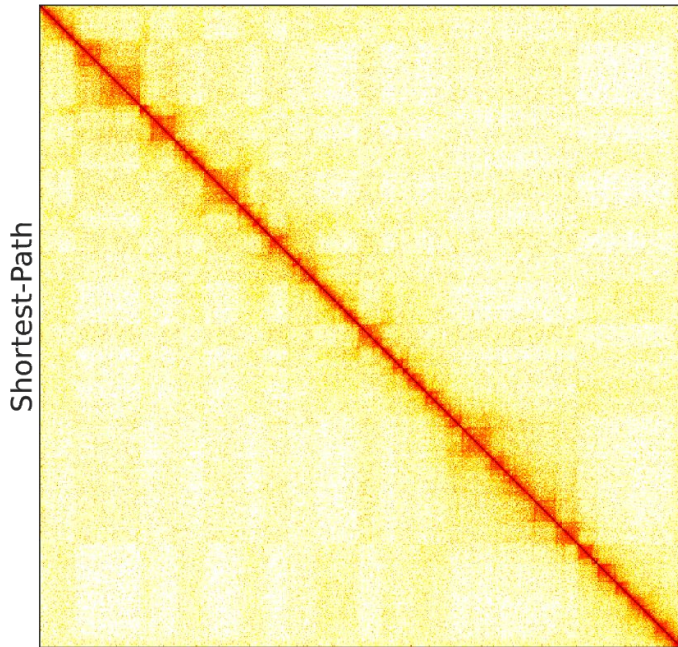
Rewired contact

Black : contact rewired by the Shortest-path

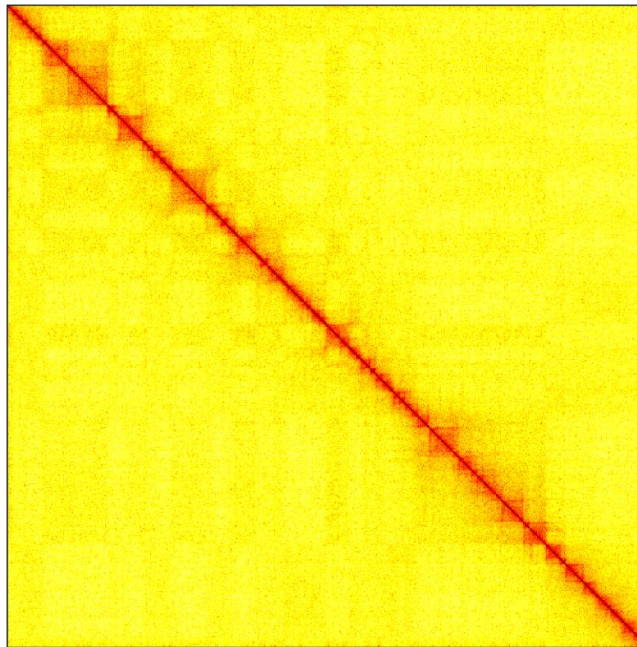


# Optimising the alpha parameter

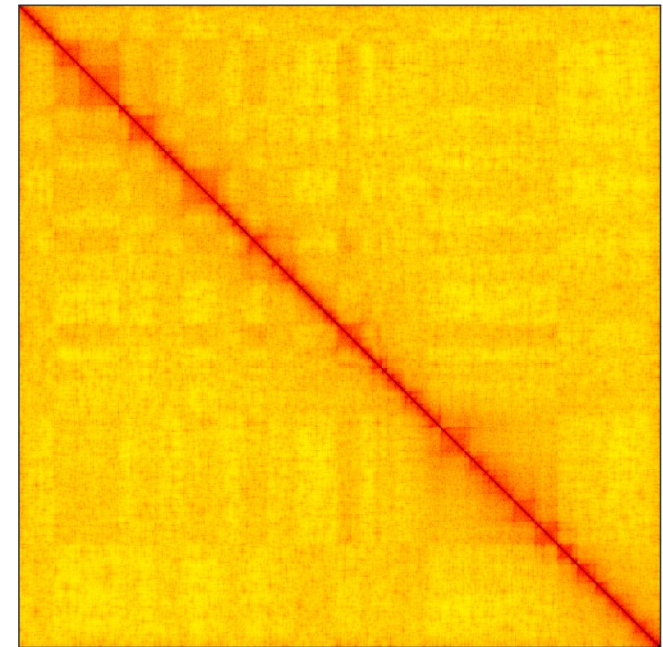
$\alpha = 0.18$



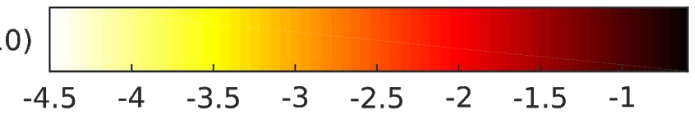
$\alpha = 0.32$



$\alpha = 0.6$

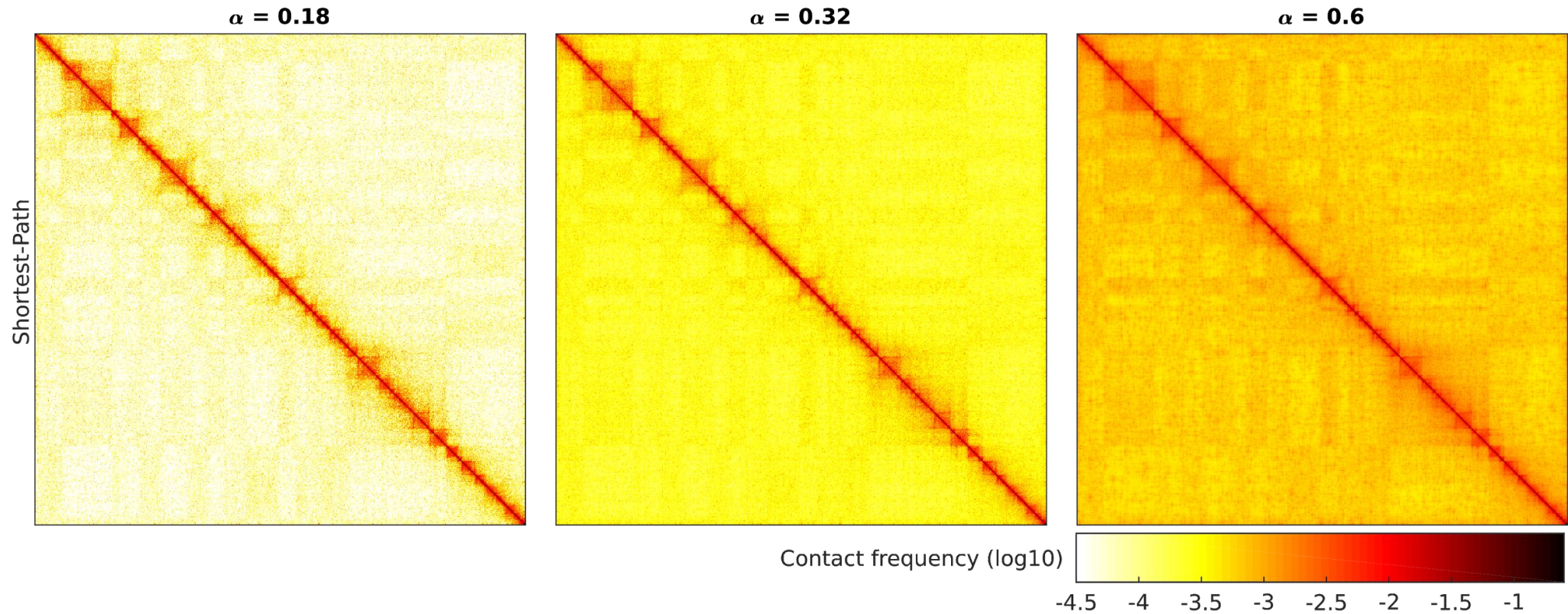


Contact frequency (log10)



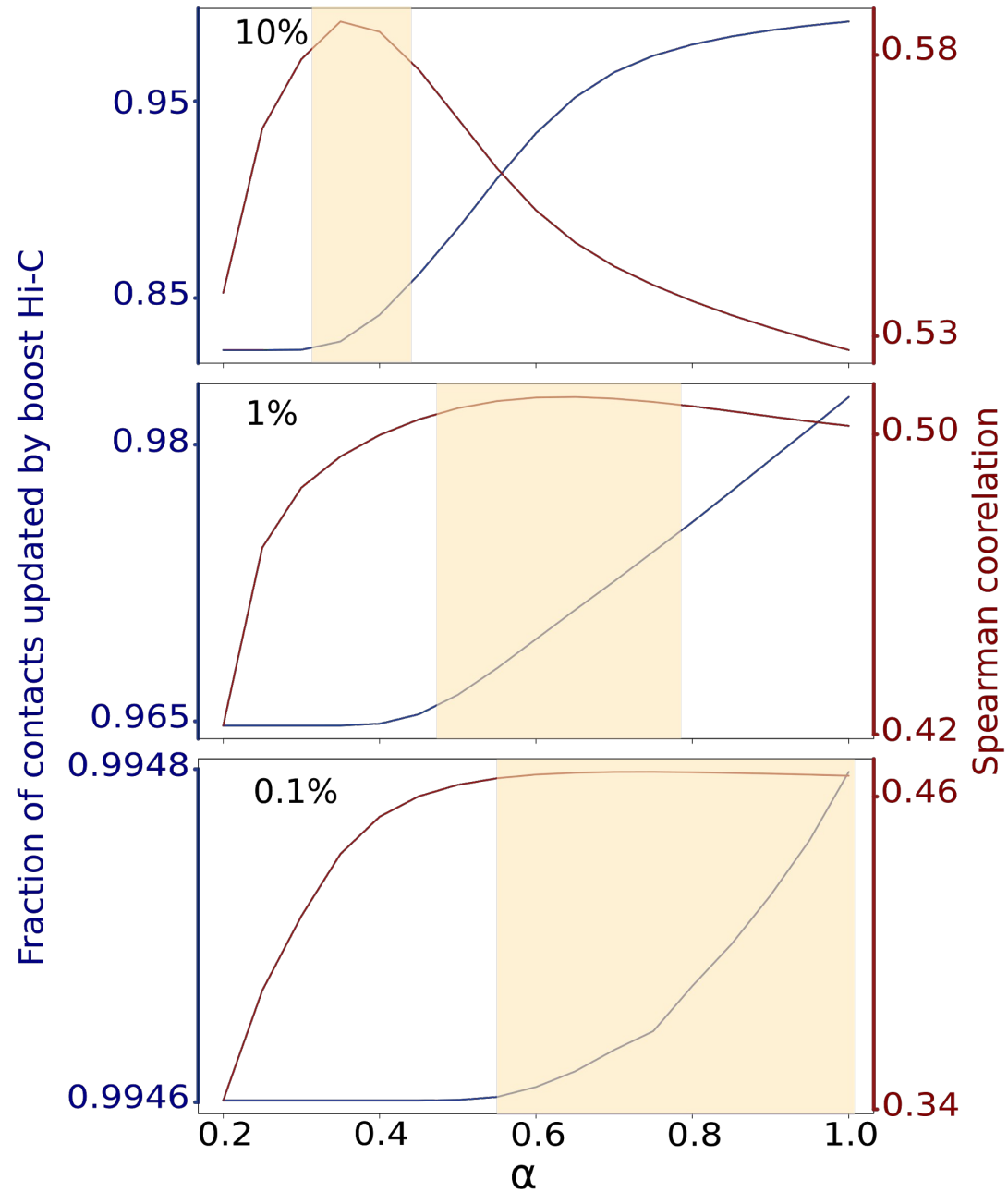


# Optimising the alpha parameter

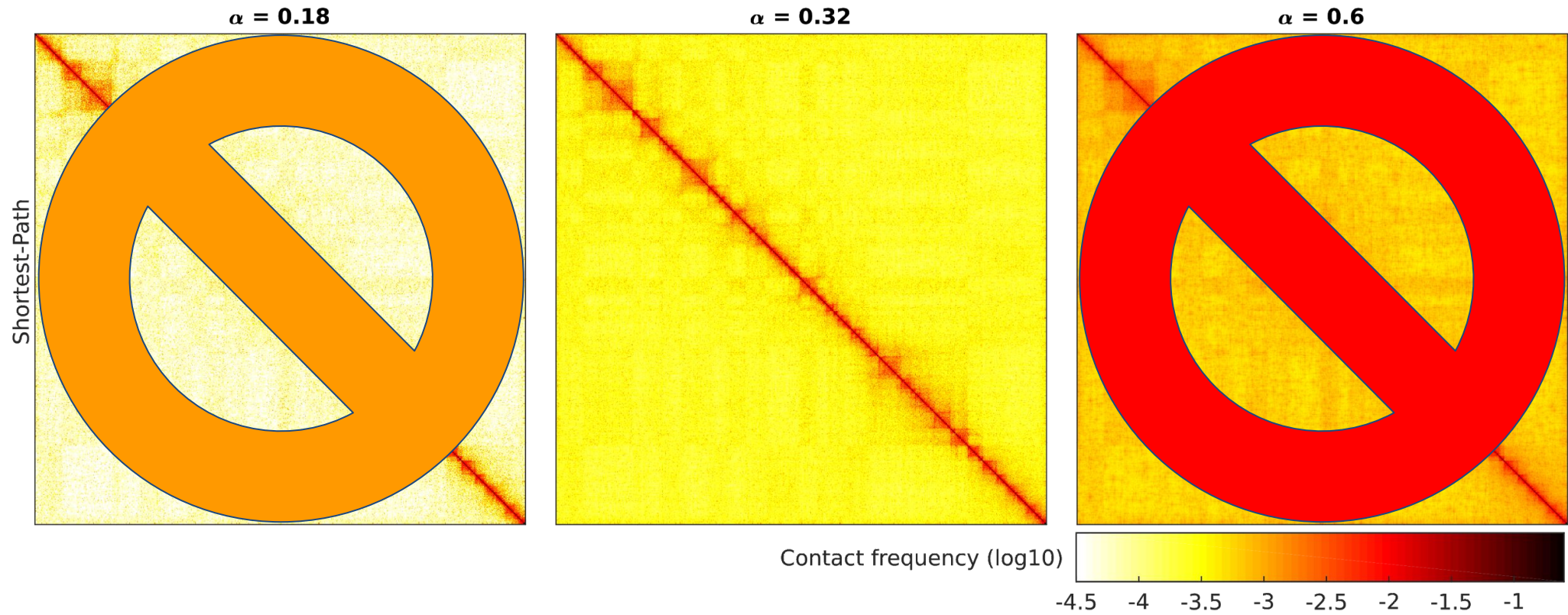


What is the optimal value of alpha?

# Optimising the alpha parameter



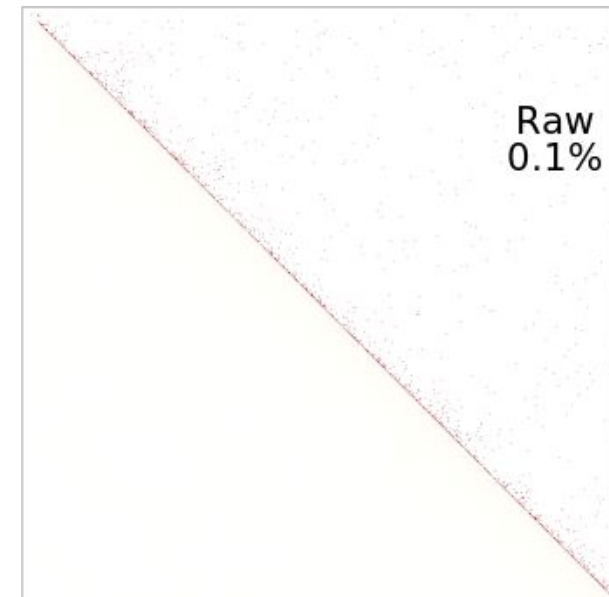
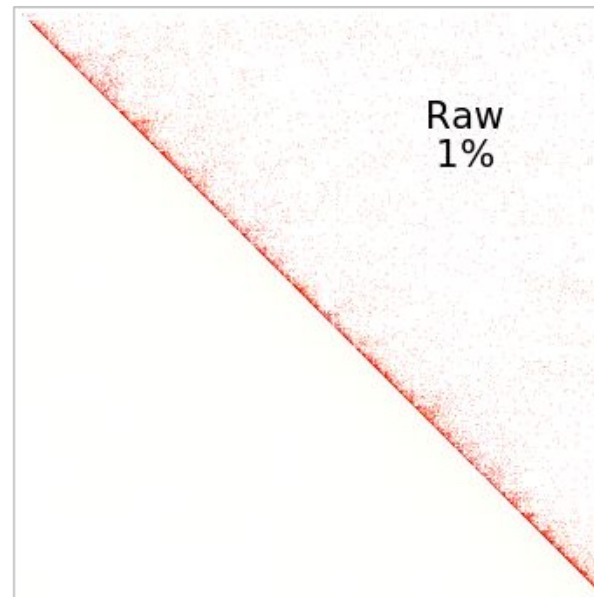
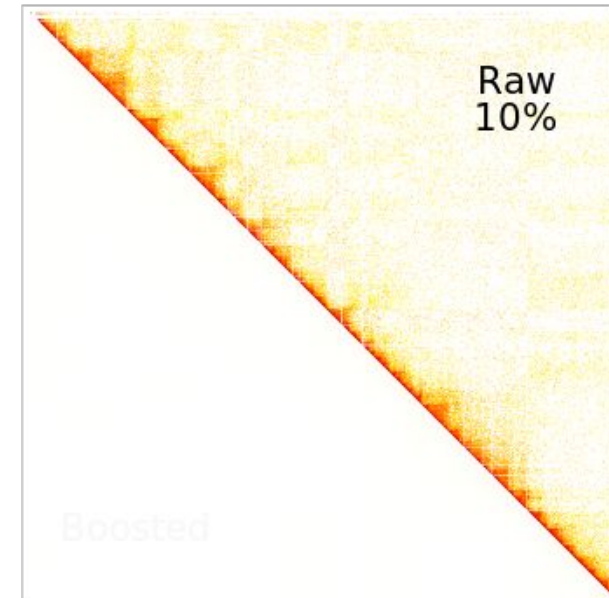
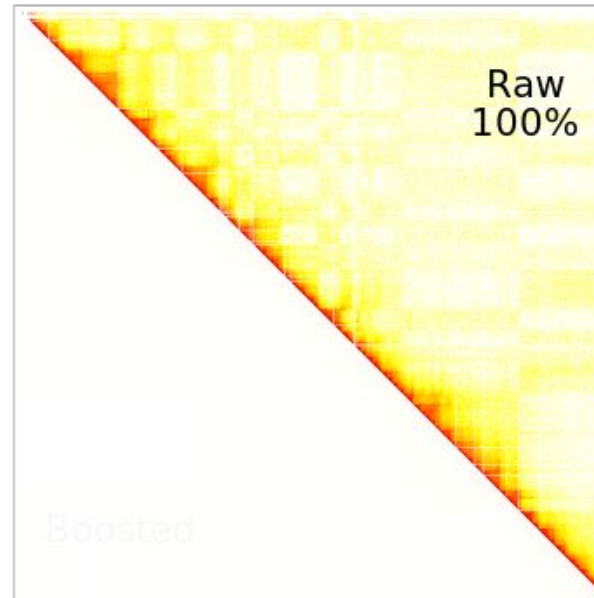
# Optimising the alpha parameter





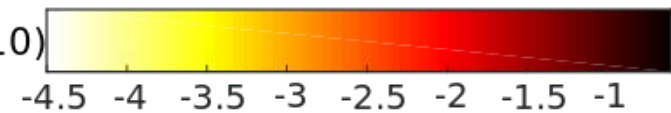
## Enhance the signal

chr16, ES-cell,  
100kb-29,8Mb



Can we impute the missing  
signal from the contact map?

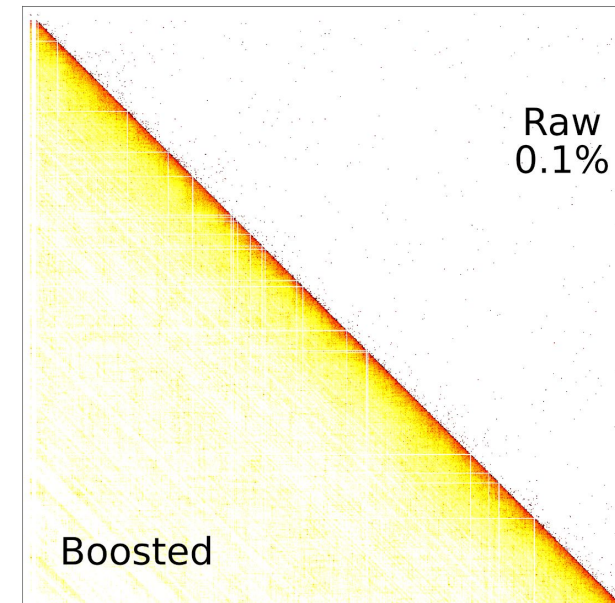
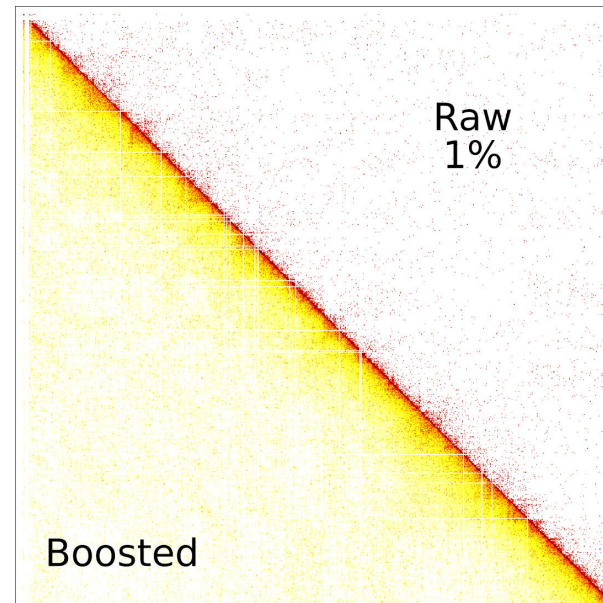
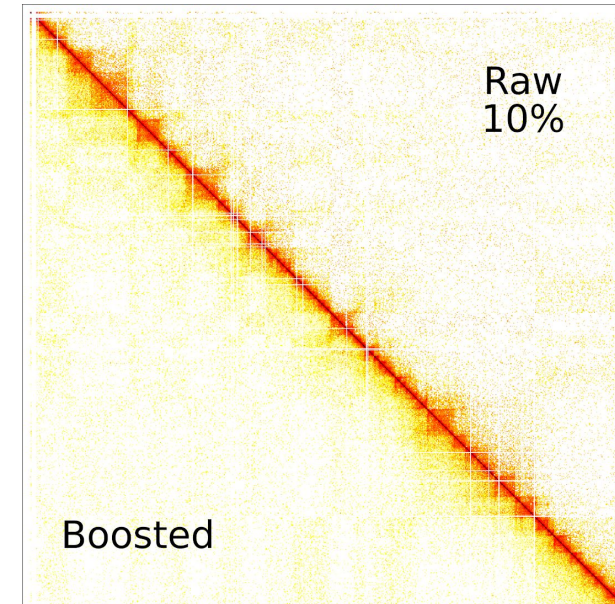
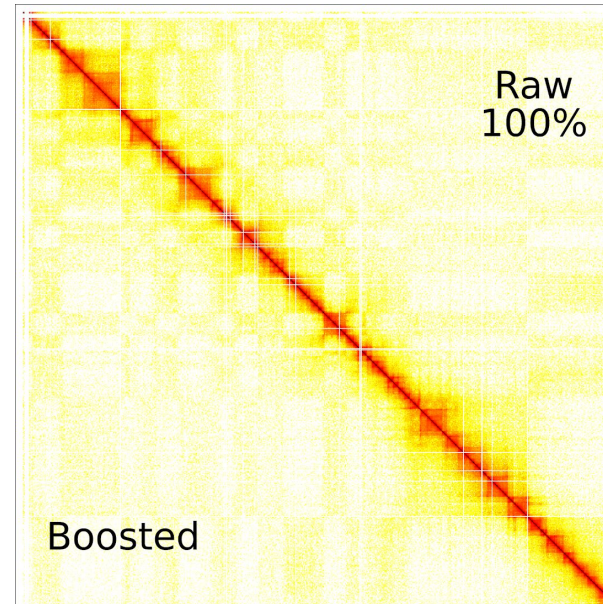
Contact frequency (log10)



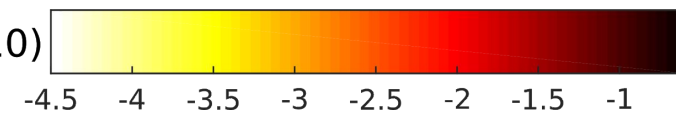
# Enhance the signal

chr16, ES-cell,  
100kb-29,8Mb

Yes we can!

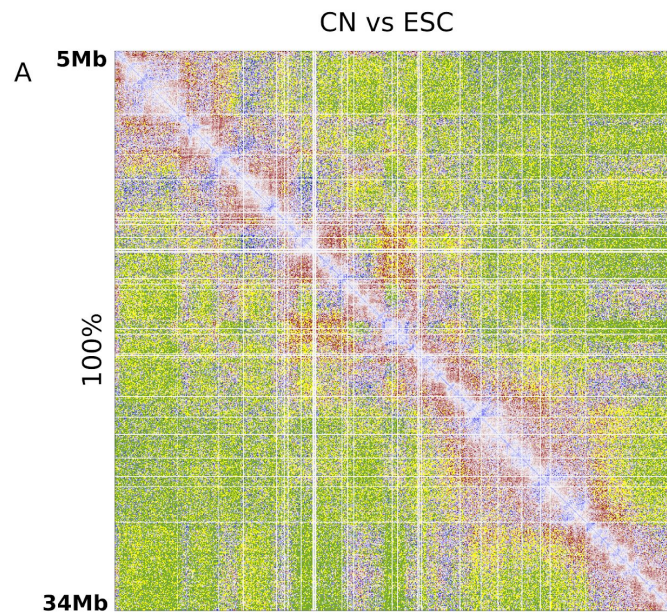


Contact frequency (log10)





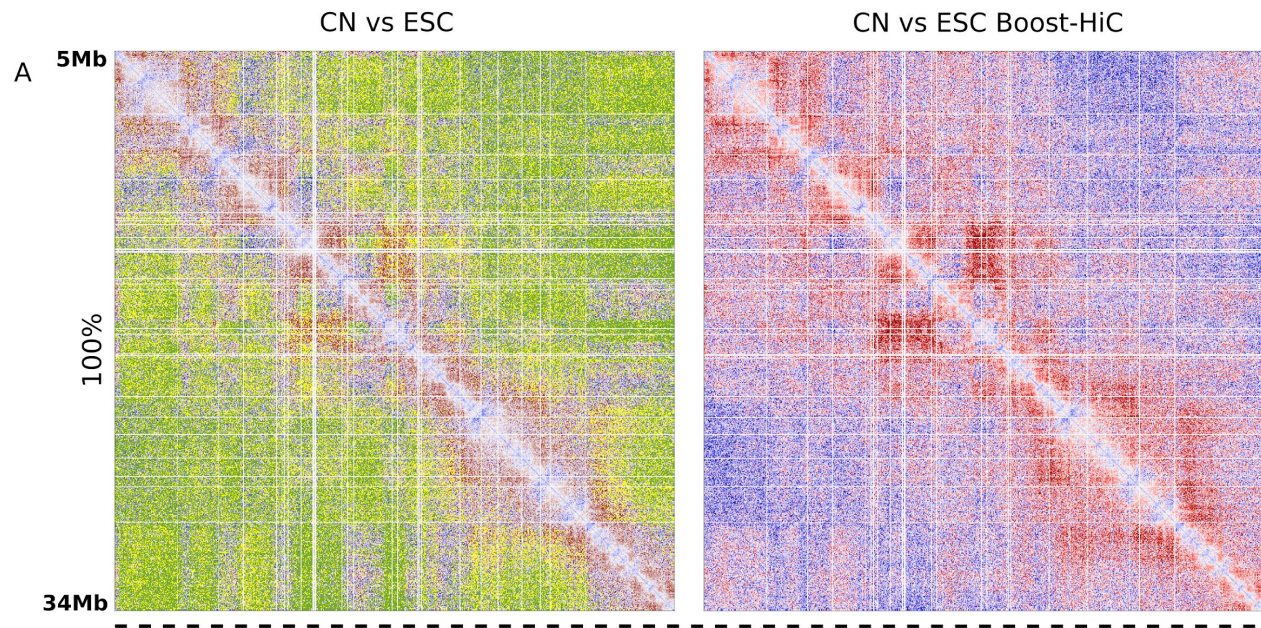
# A real improvement : log ratio



Contact ratio (log2)

-inf -3 -2 -1 0 1 2 3 +inf

# A real improvement : log ratio

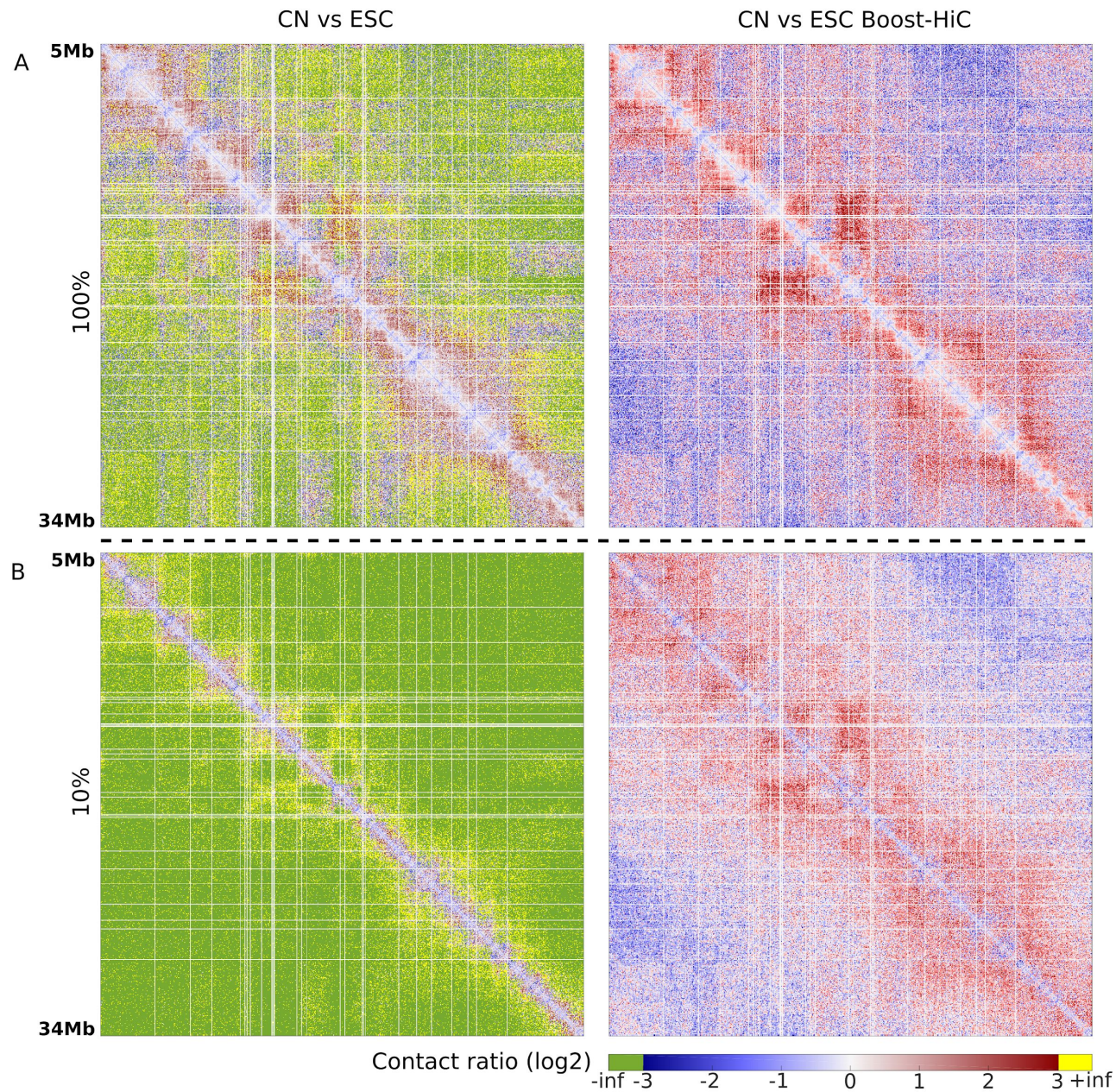


Contact ratio (log2)

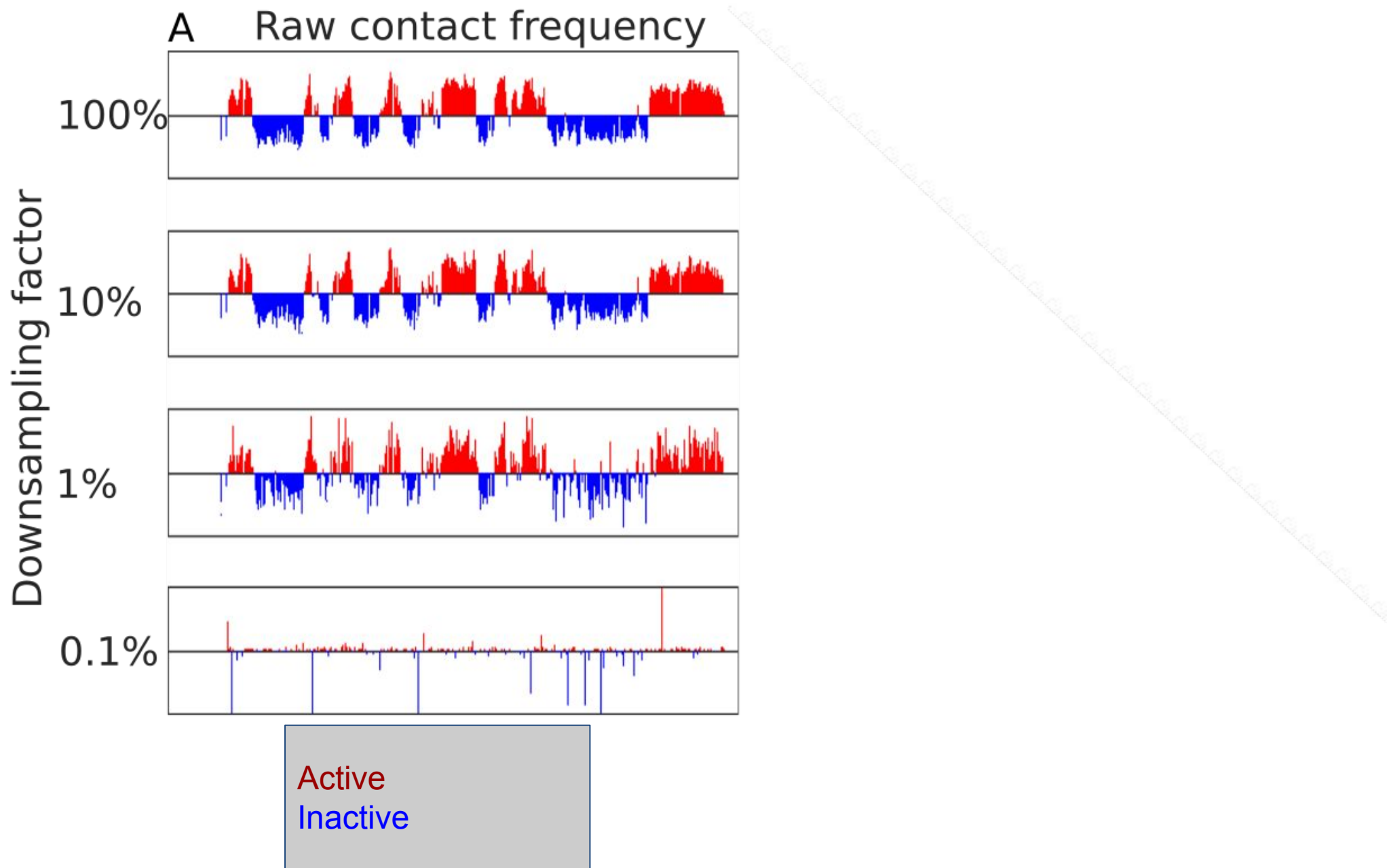
-inf -3 -2 -1 0 1 2 3 +inf



# A real improvement : log ratio

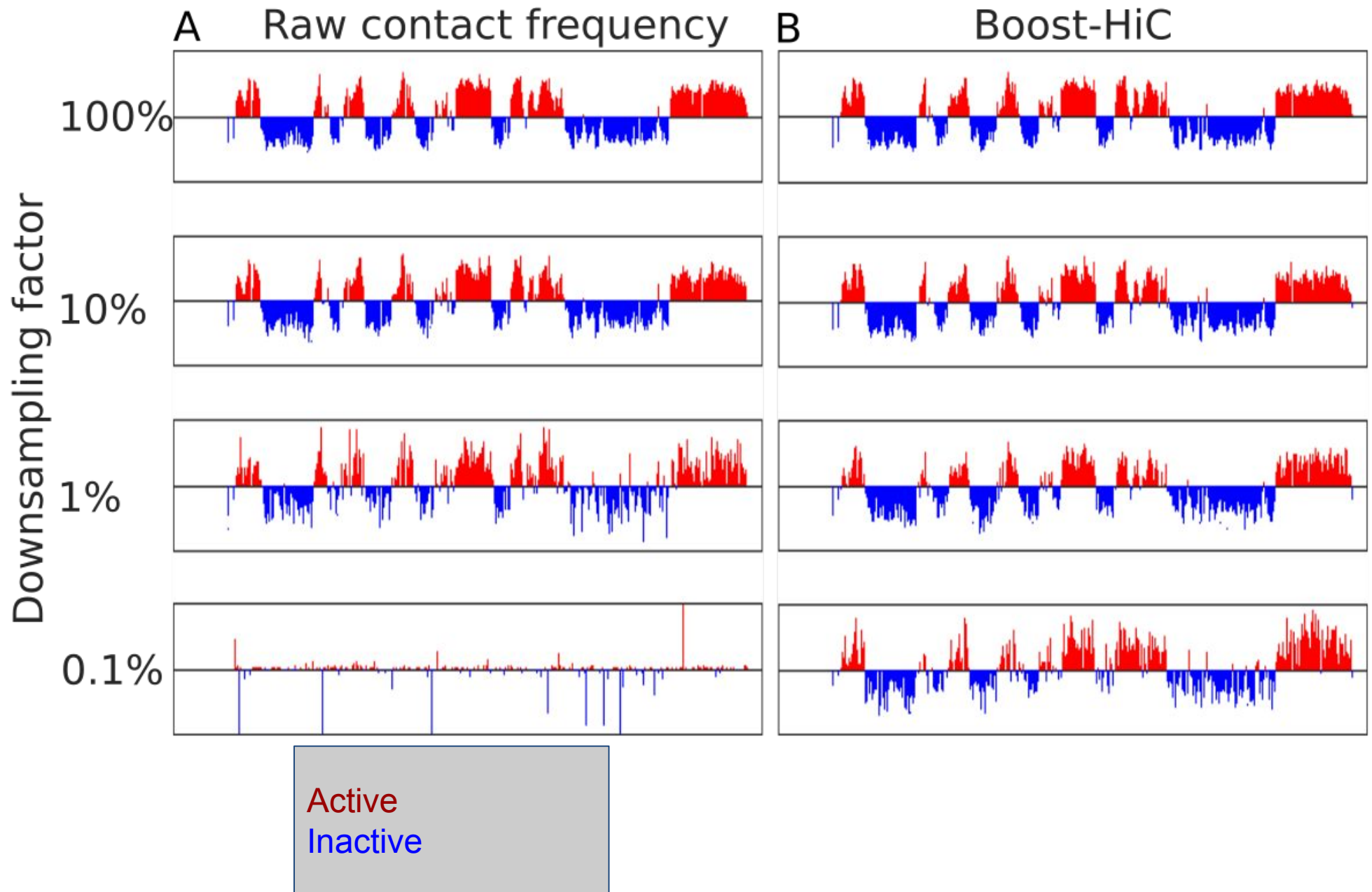


# Boost-HiC improve genomic compartment for low sequencing dataset

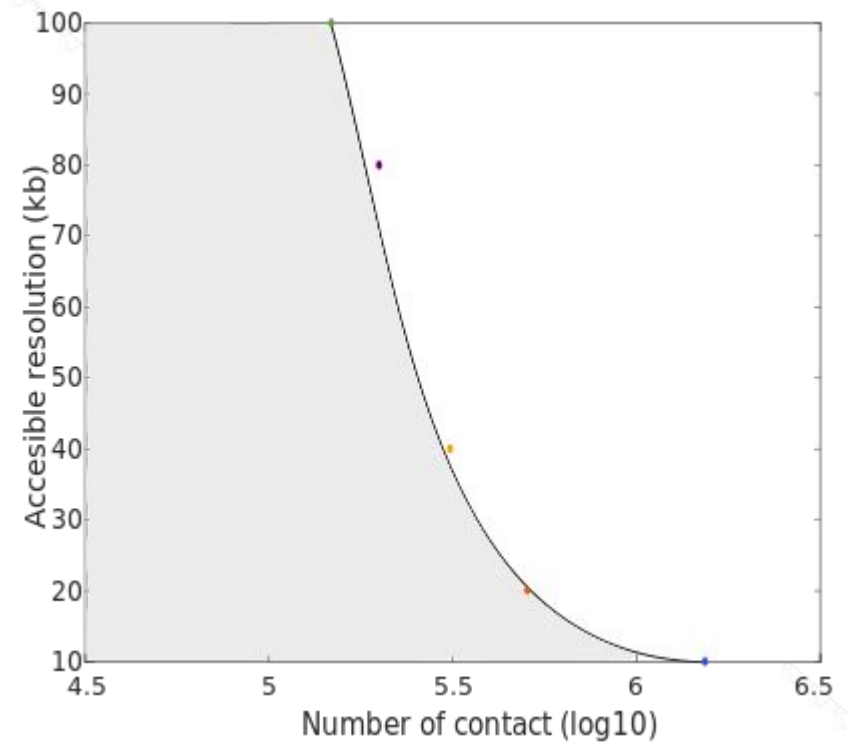
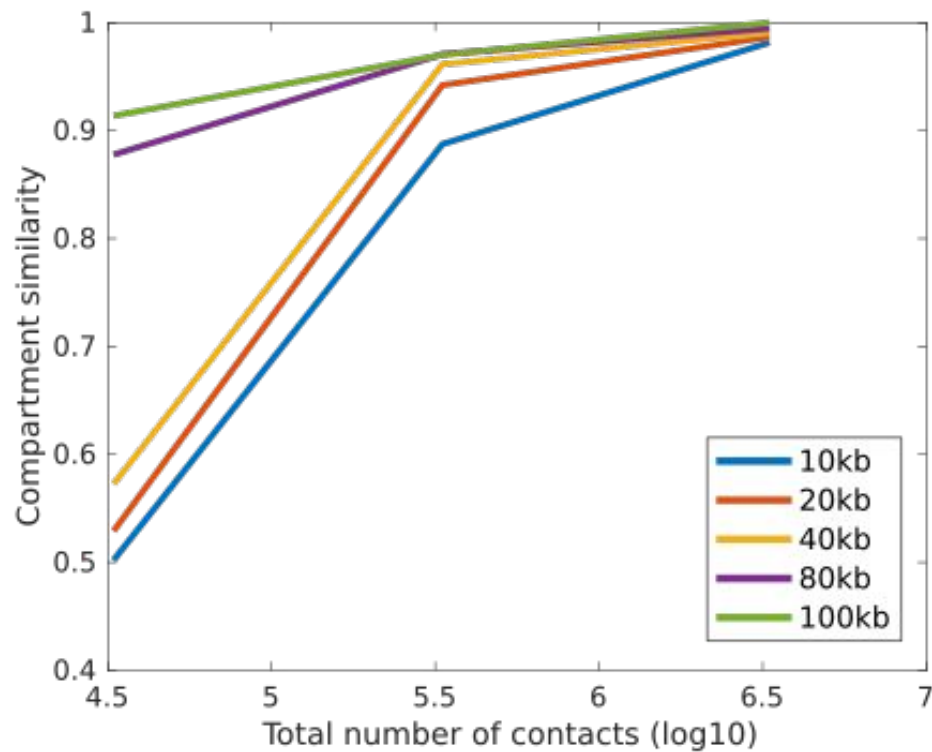




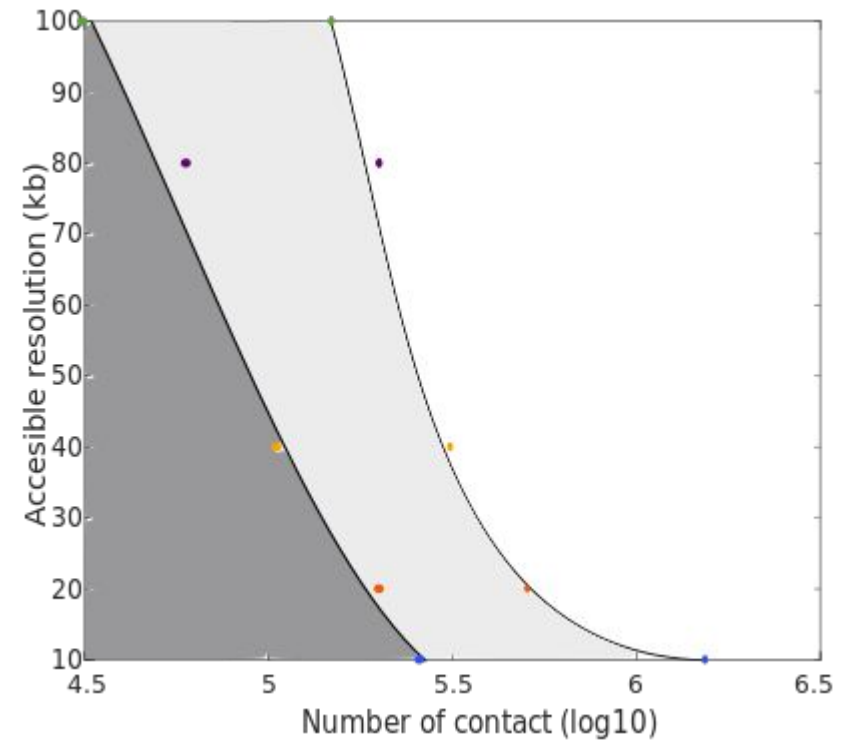
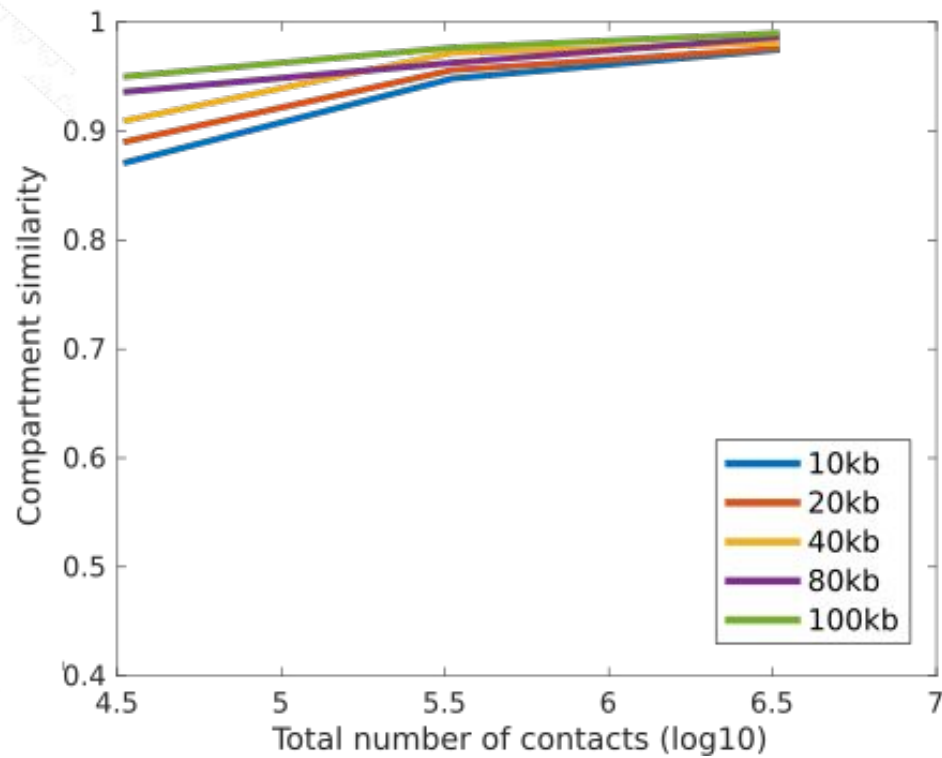
# Boost-HiC improve genomic compartment for low sequencing dataset



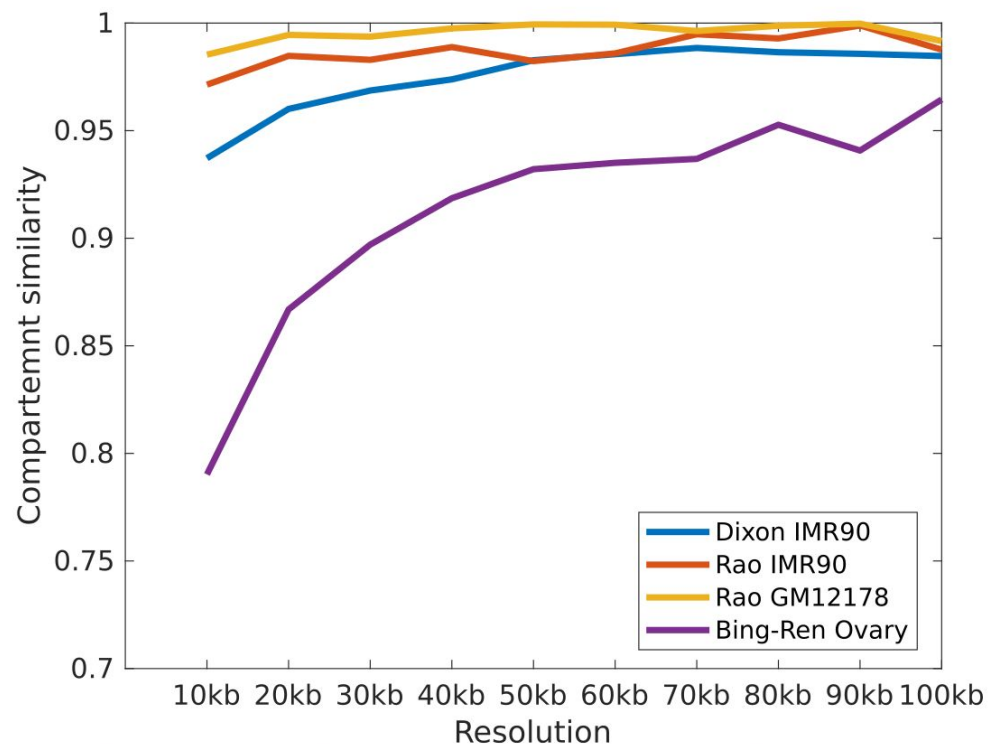
# Optimal sequencing : Raw information



# Optimal sequencing : Raw information



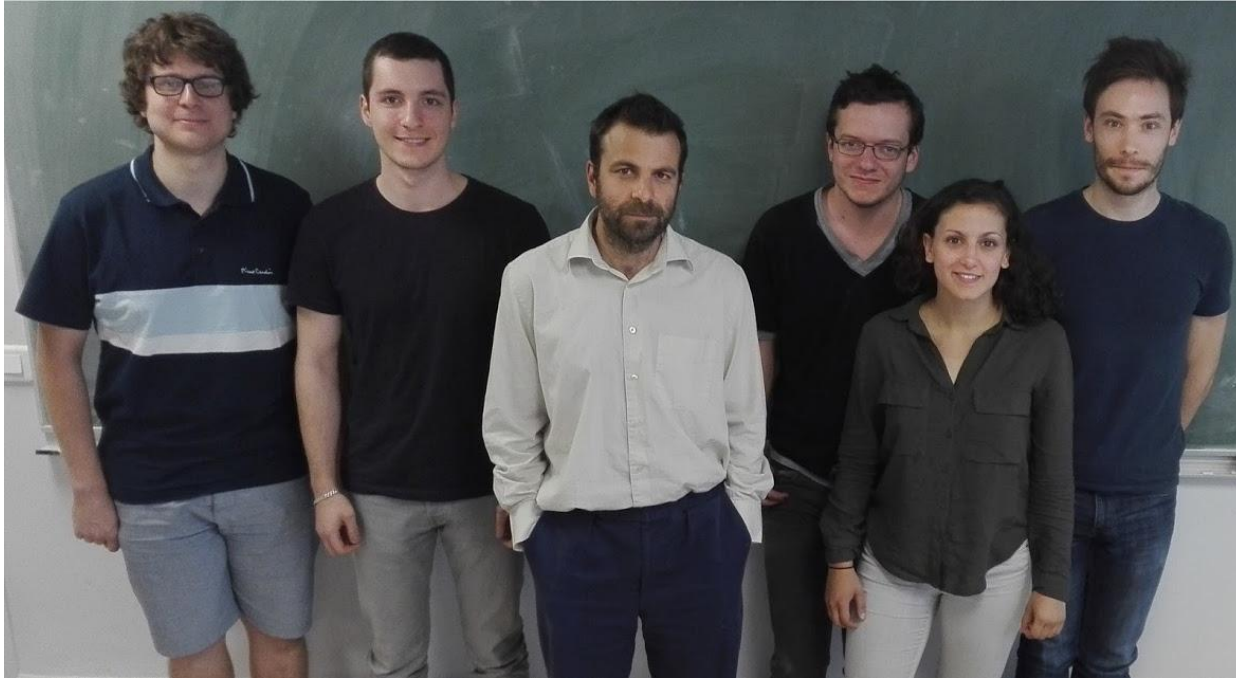
# Optimal sequencing : Other information on raw dataset



# Conclusions

- + Boost HiC is a good tools to study genomic compartment at high resolution and low sequencing condition
- + You don't always need to have high sequencing step

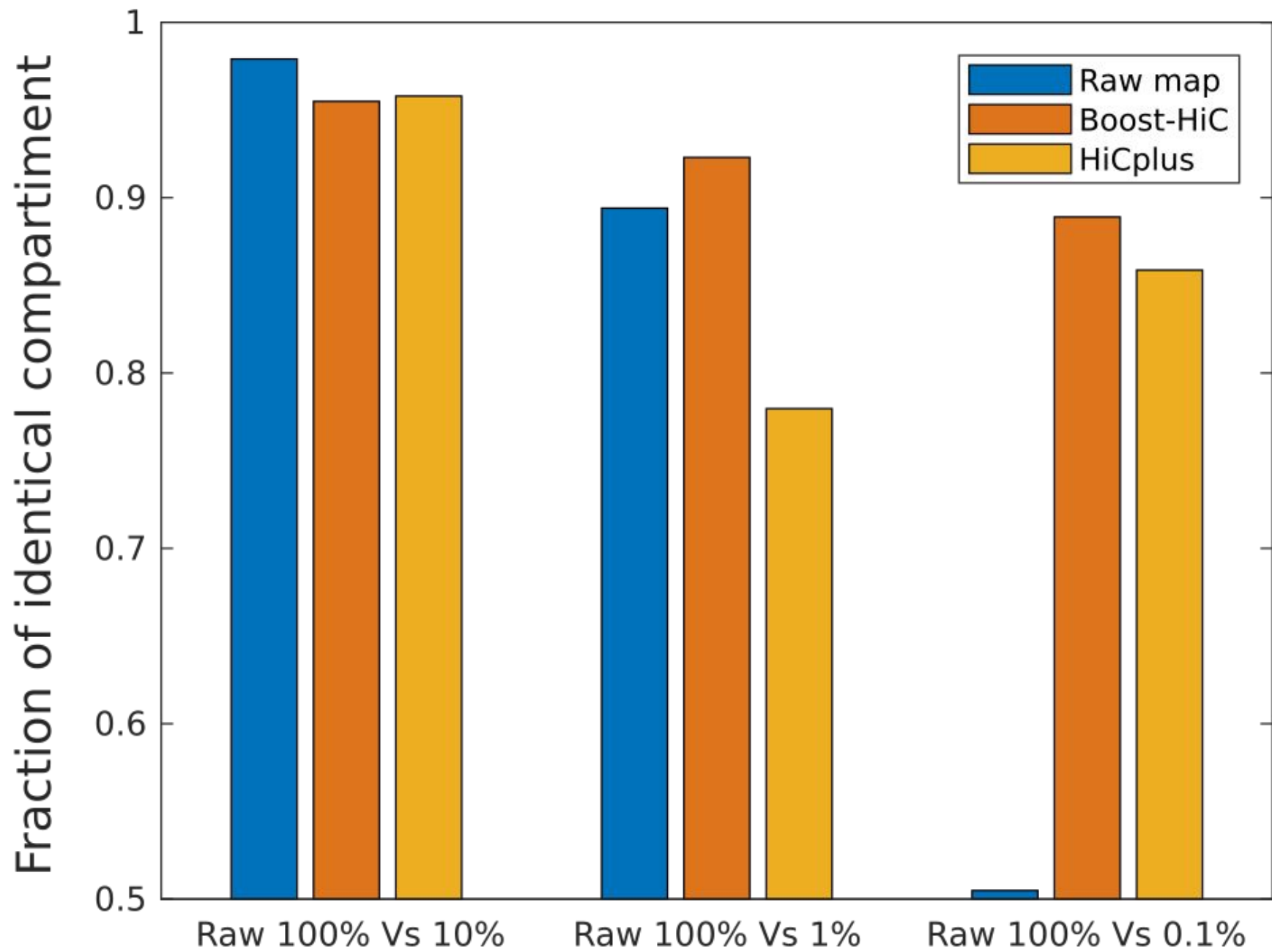
# Acknowledgment



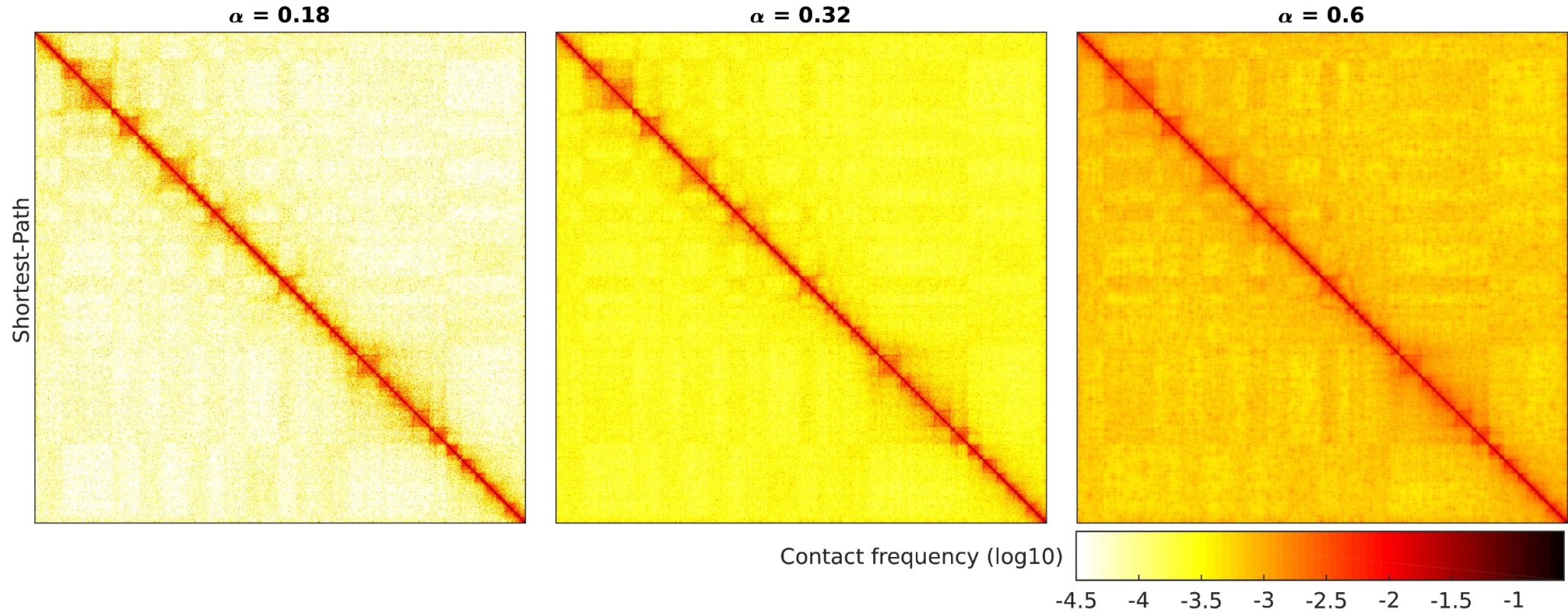
Julien Mozziconacci  
Jean Baptiste Morlot  
Vincent Matthys  
Annick Lesne  
People from LPTMC



## Some advantage of the procedure

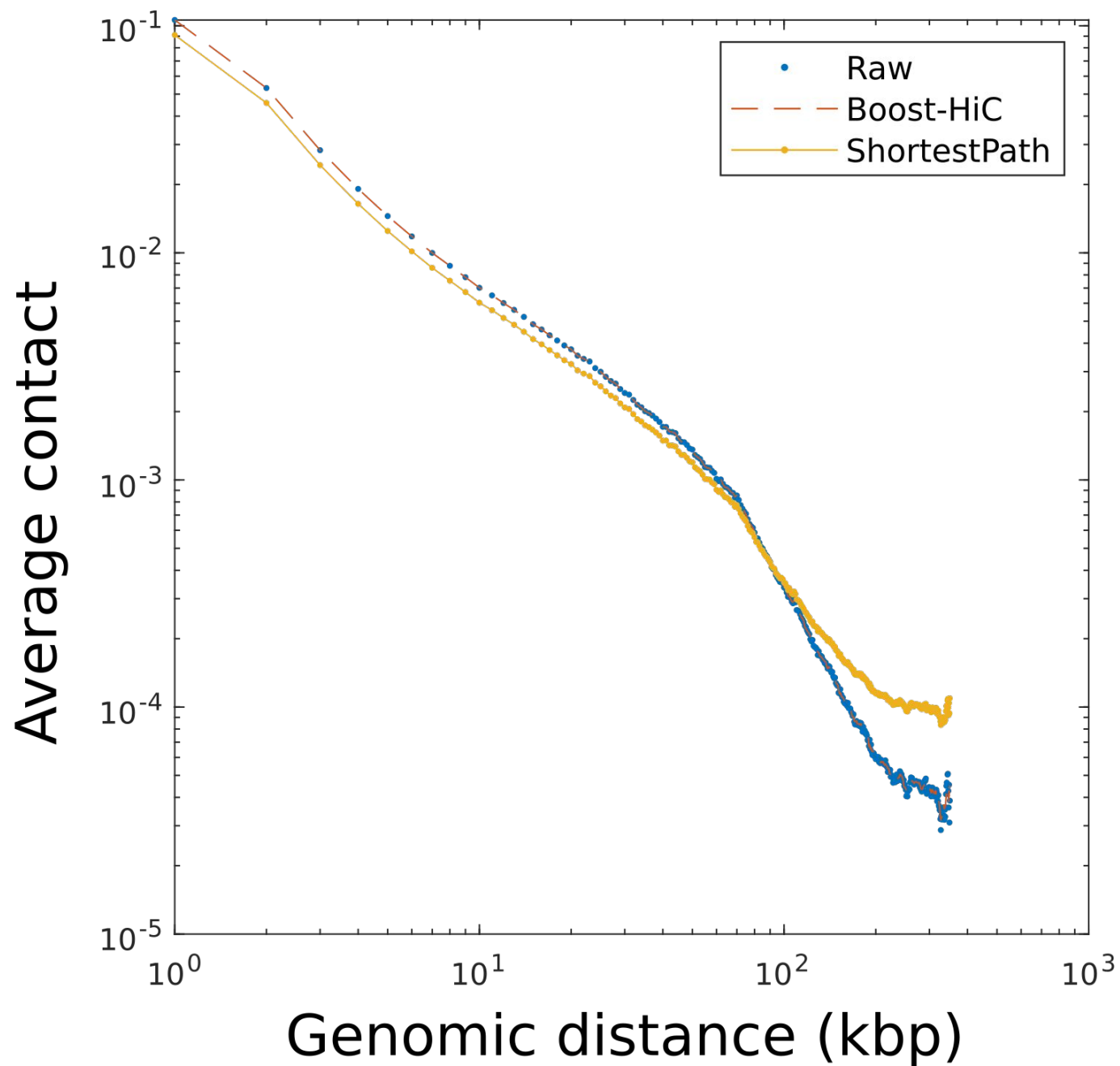


# Optimising the alpha parameter

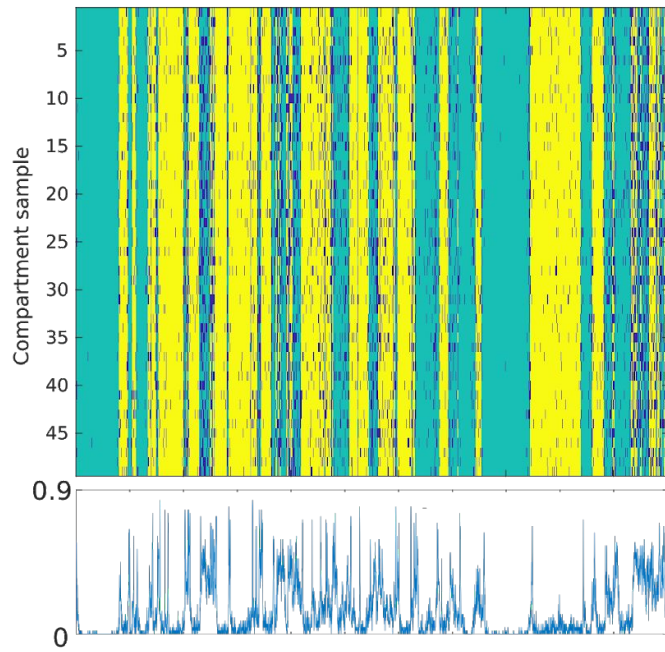


Did we change the contact probability?

## Restore the contact probability

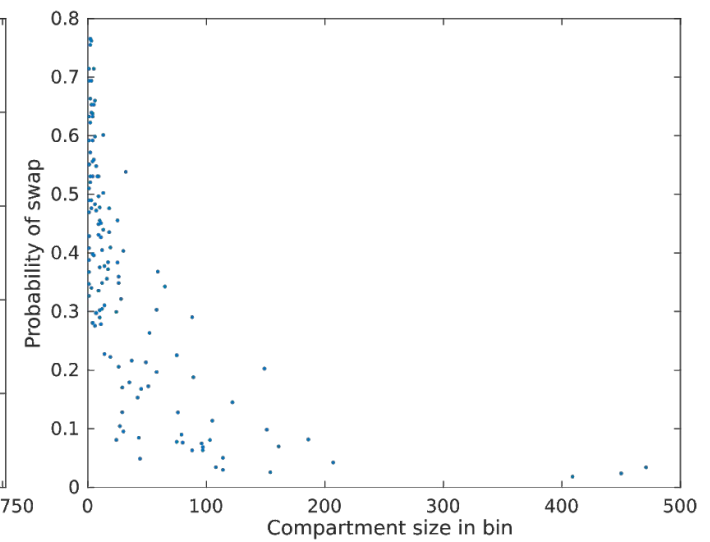
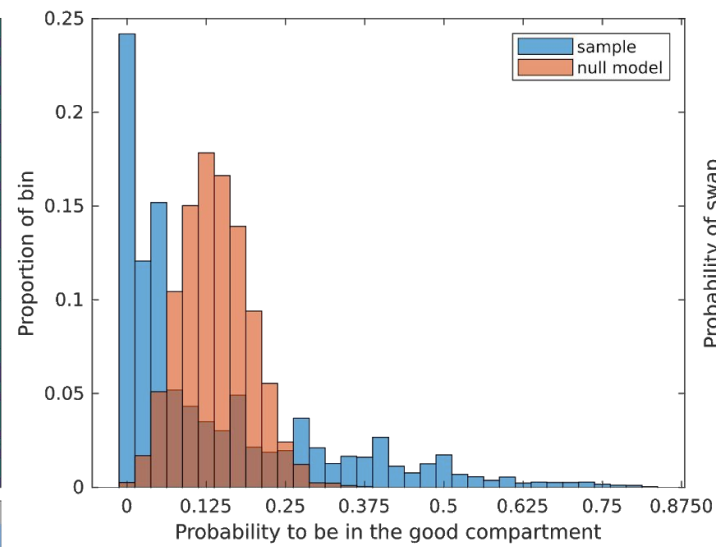
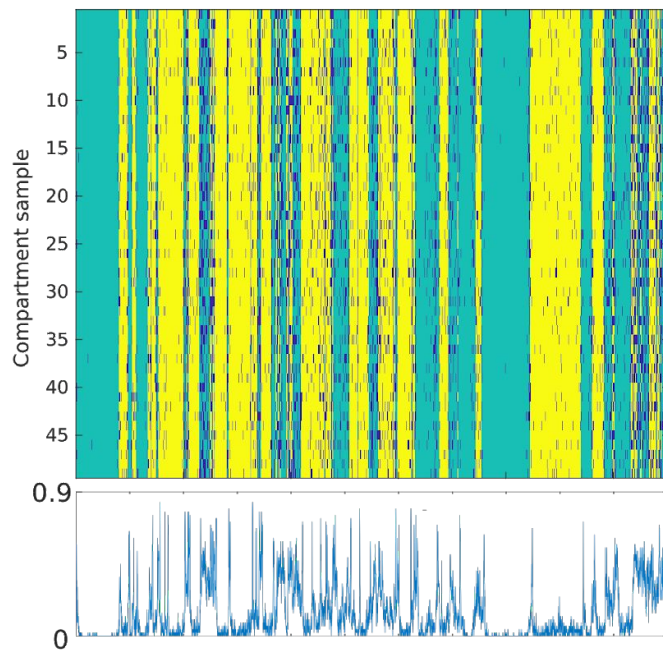


# Compartment signature



Active  
Inactive  
Swap

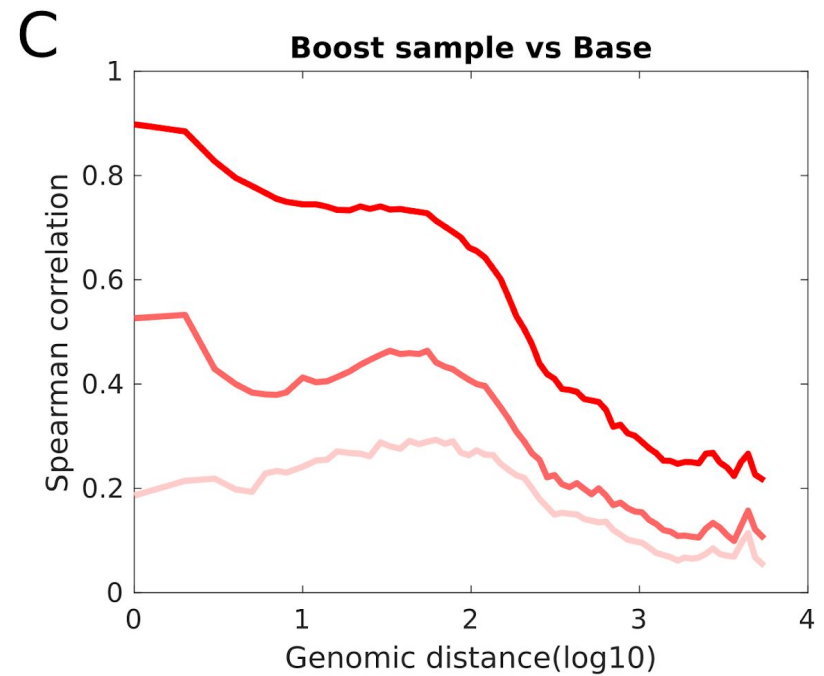
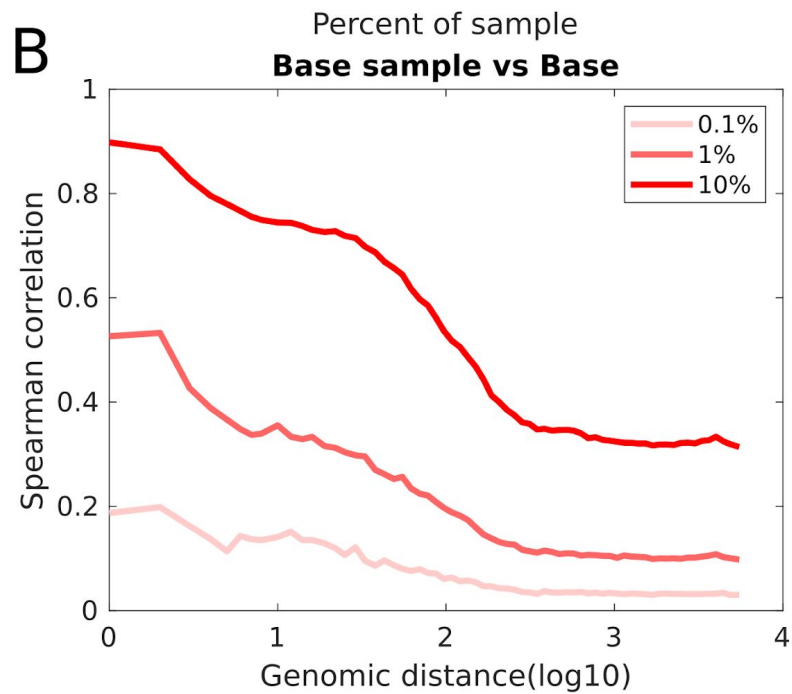
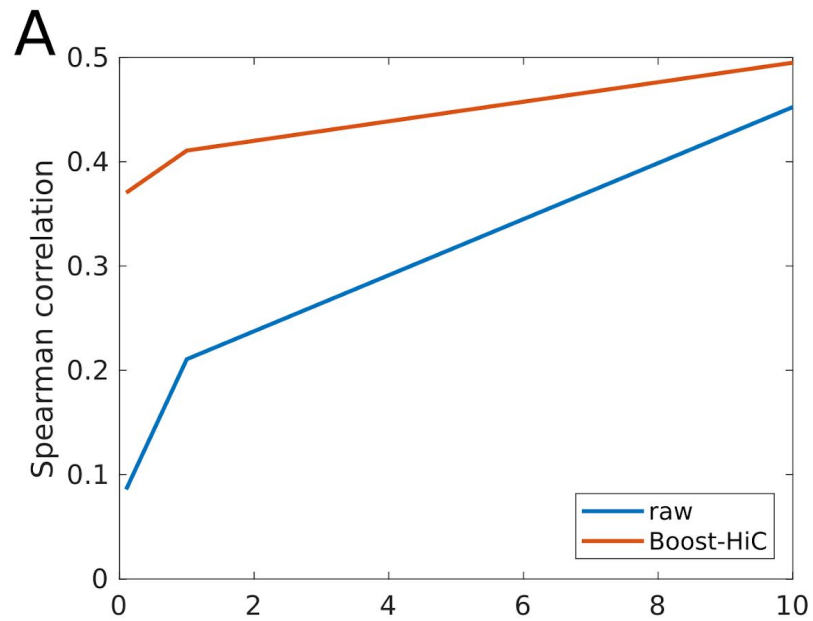
# Compartment signature



Active  
Inactive  
Swap



# Effect on distance genomic



# Number of reads, contact probability

