

3DAROC16

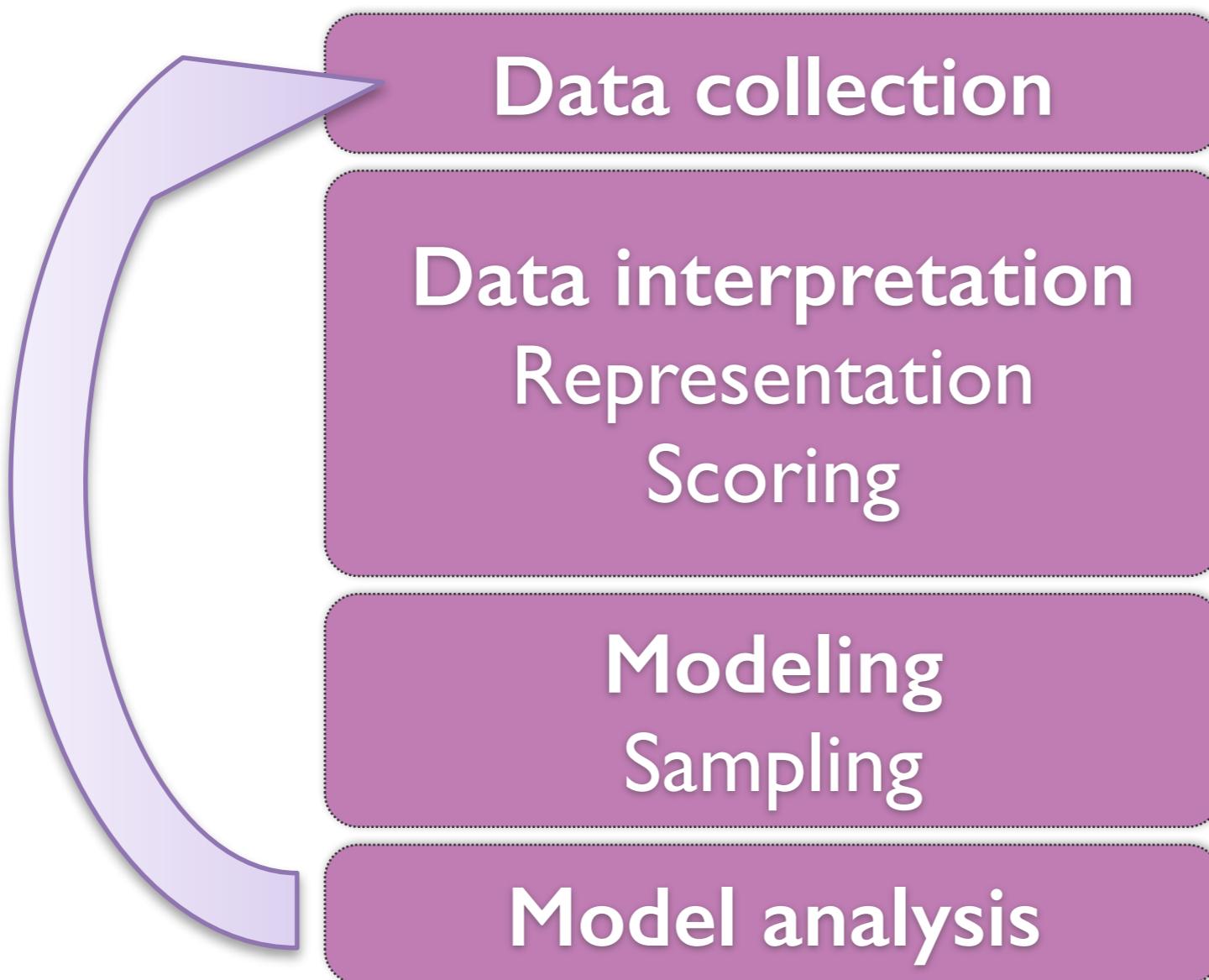
Summary day #3

**Marco Di Stefano, François Serra &
Marc A. Marti-Renom**
Structural Genomics Group (CNAG-CRG)

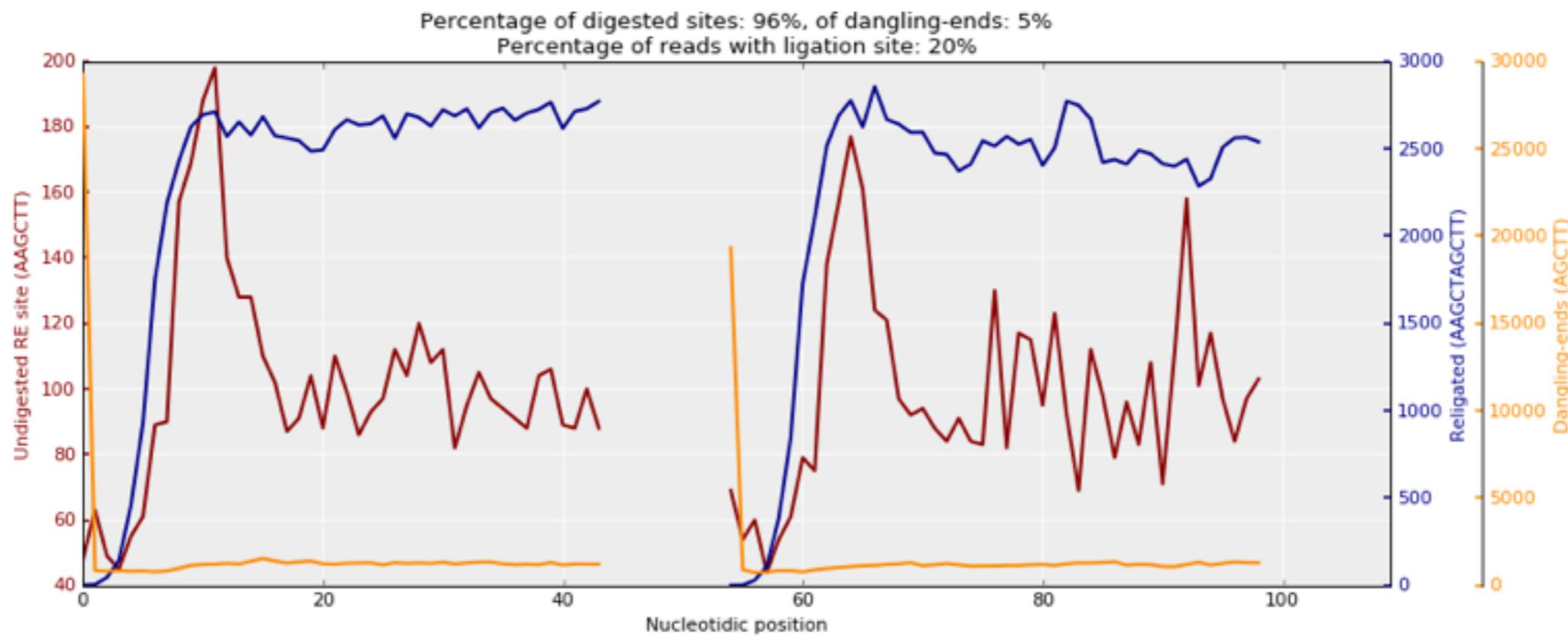
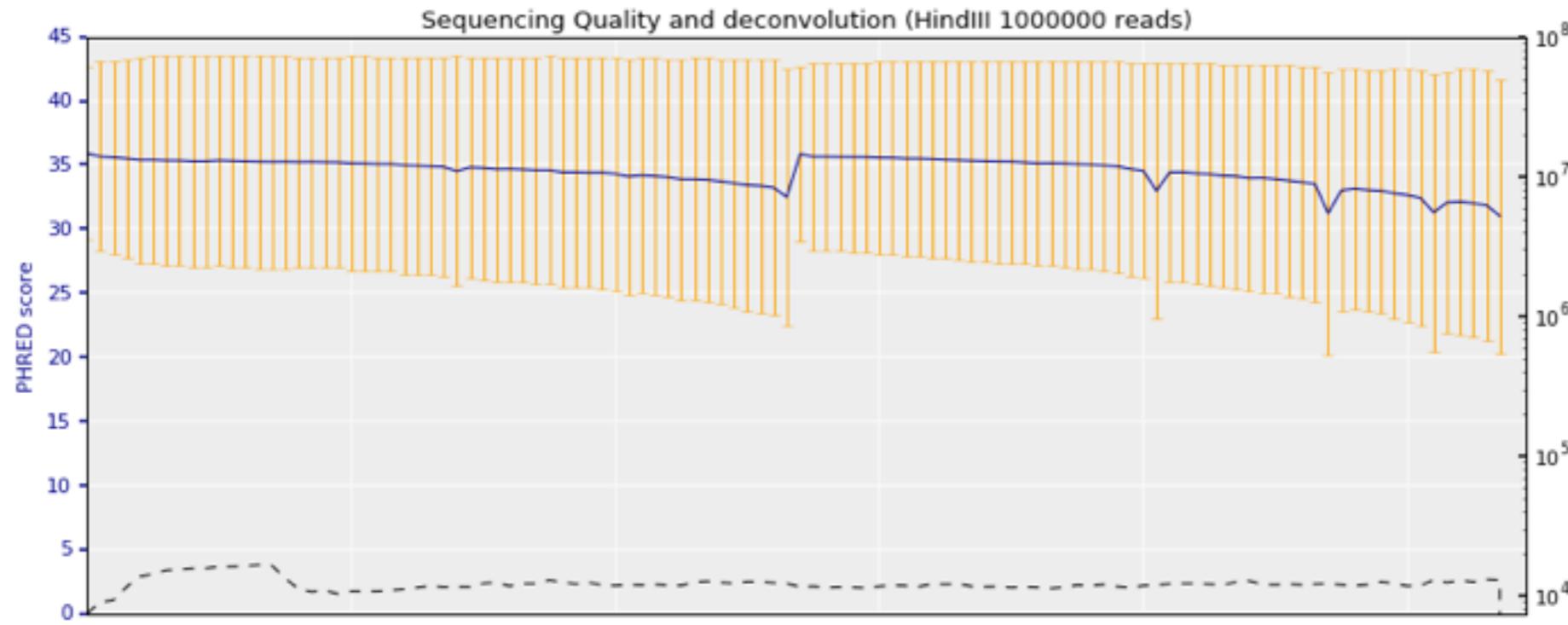


Day 1

Integrative modeling



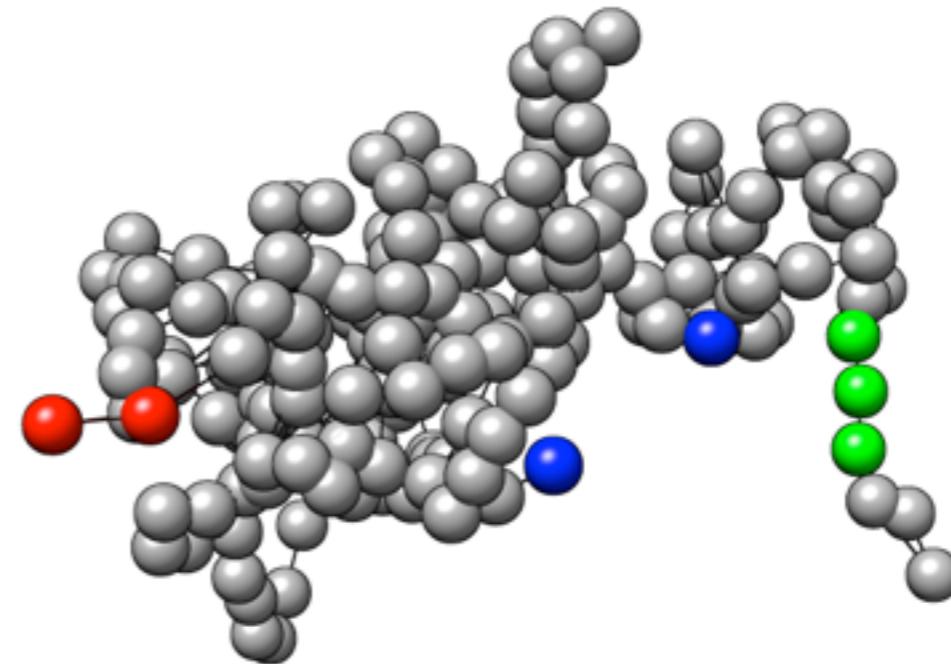
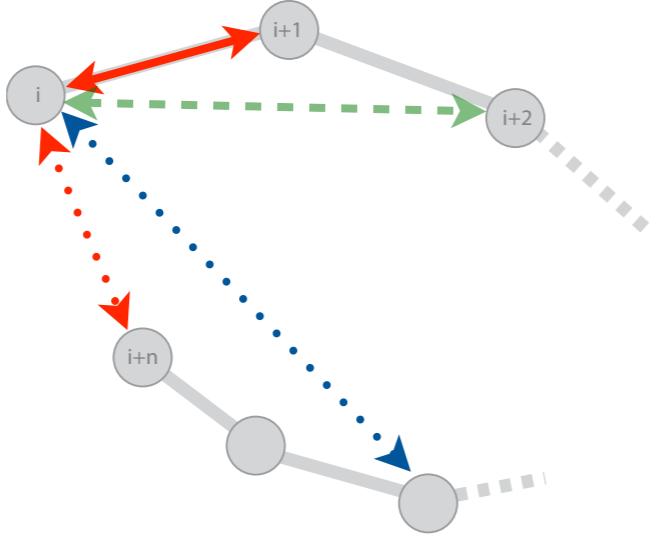
Quality plots of the reads



Day 2

Model representation and scoring

Constituent parts of the molecule



$$d < d_0$$



$$d = d_0$$

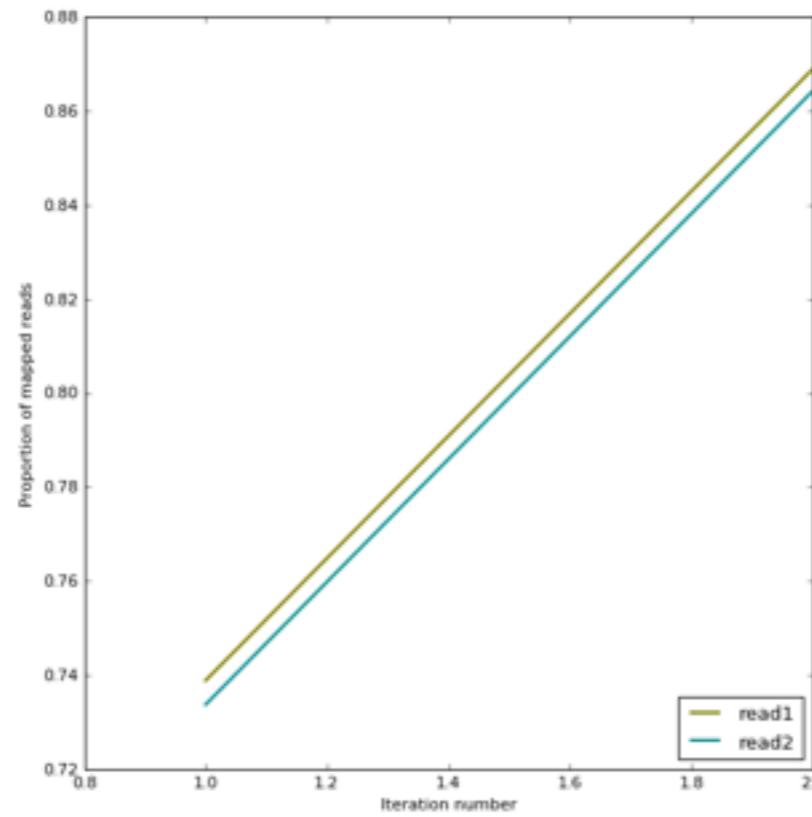


$$d > d_0$$

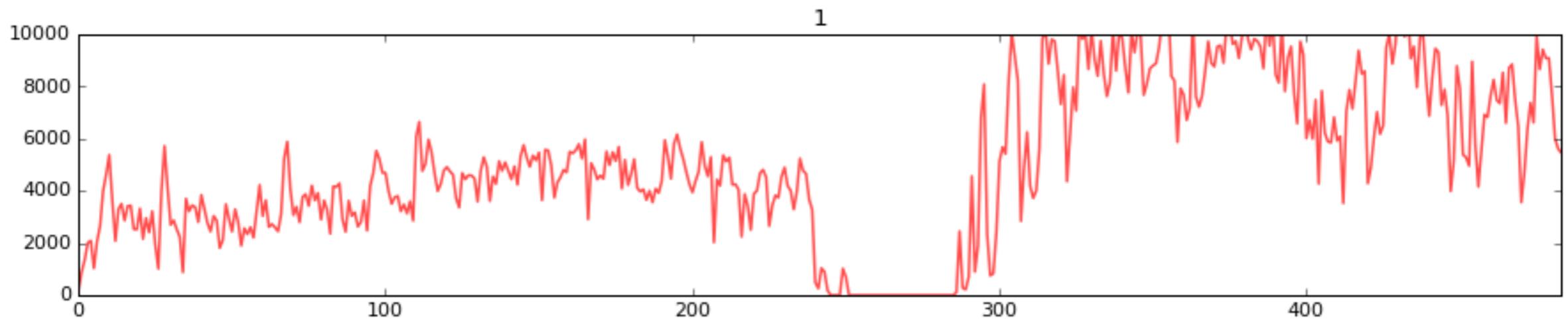
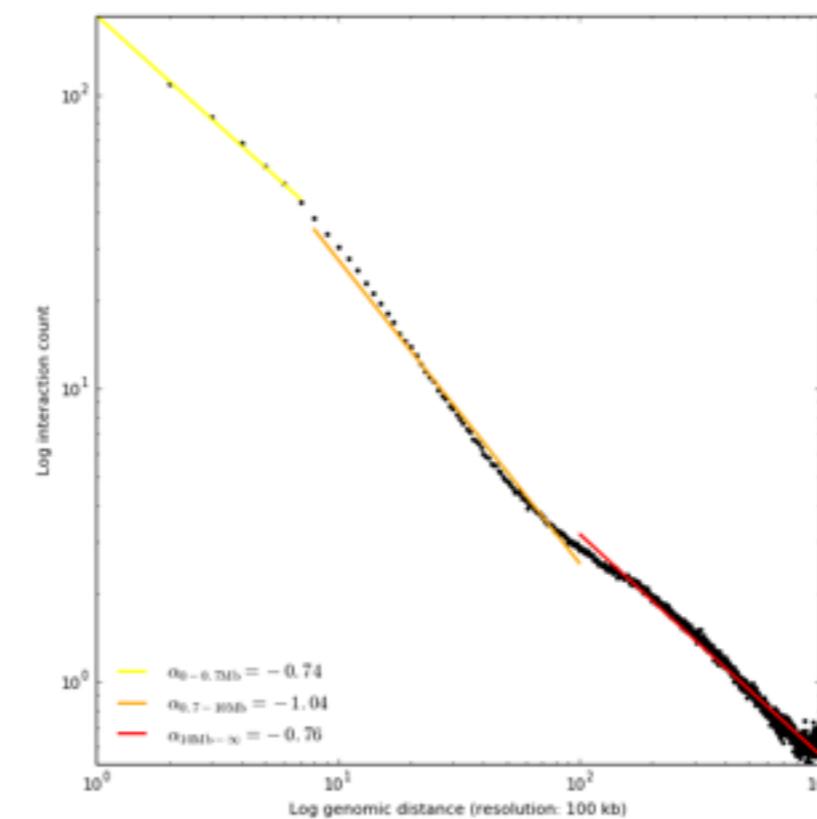


Data quality check

Mapped reads vs full/frag



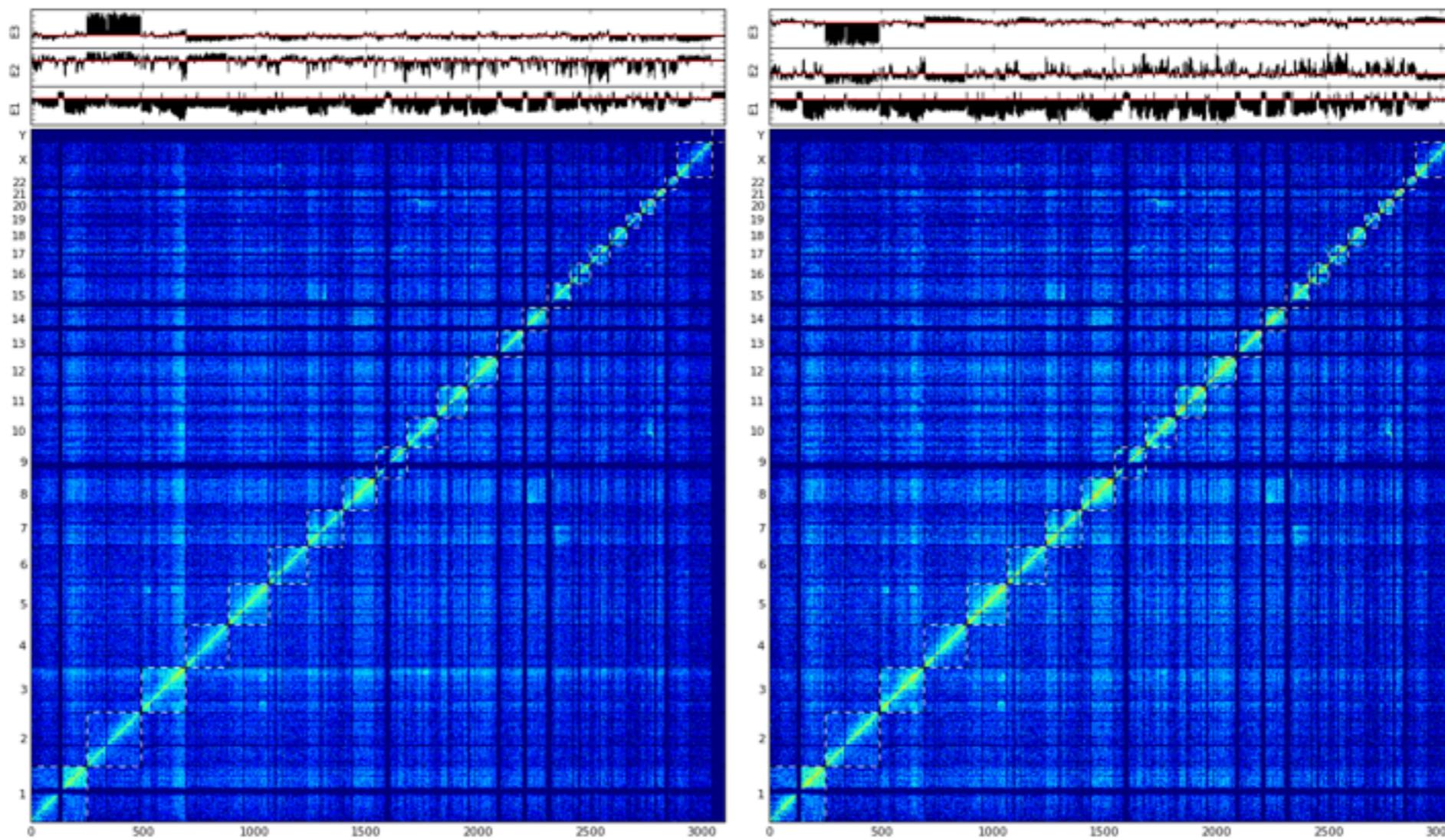
Interactions vs genomic distance



Hi-C map generation and filtering

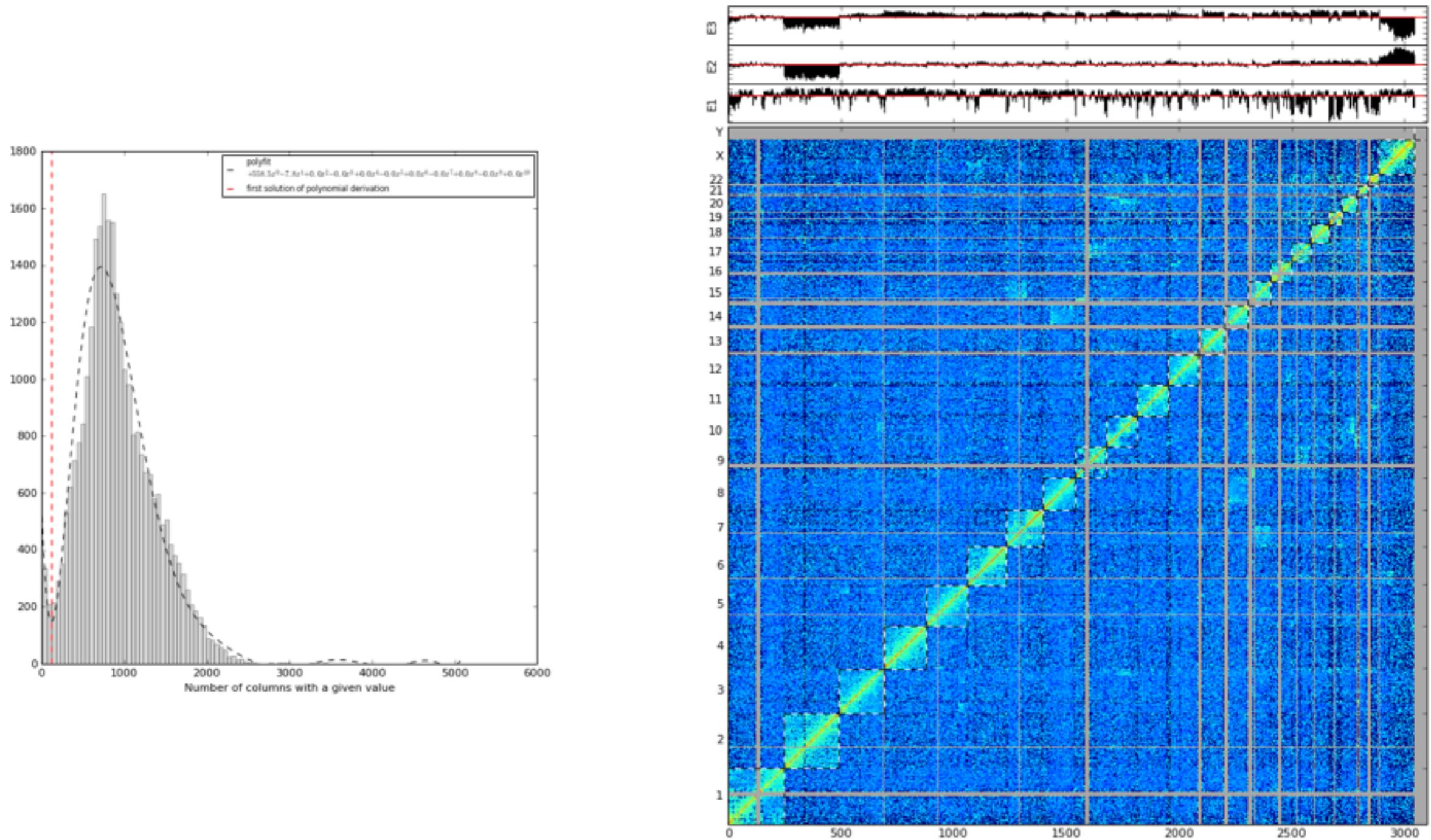
Filtered reads (and percentage of total):

| Mapped both | : | 21258763 (100.00%) |
|-----------------------|---|--------------------|
| 1- self-circle | : | 400576 (1.88%) |
| 2- dangling-end | : | 1174450 (5.52%) |
| 3- error | : | 26229 (0.12%) |
| 4- extra dangling-end | : | 2971724 (13.98%) |
| 5- too close from RES | : | 2959353 (13.92%) |
| 6- too short | : | 227026 (1.07%) |
| 7- too large | : | 2322 (0.01%) |
| 8- over-represented | : | 693664 (3.26%) |
| 9- duplicated | : | 2123831 (9.99%) |
| 10- random breaks | : | 2342747 (11.02%) |



Hi-C map filtering & normalising

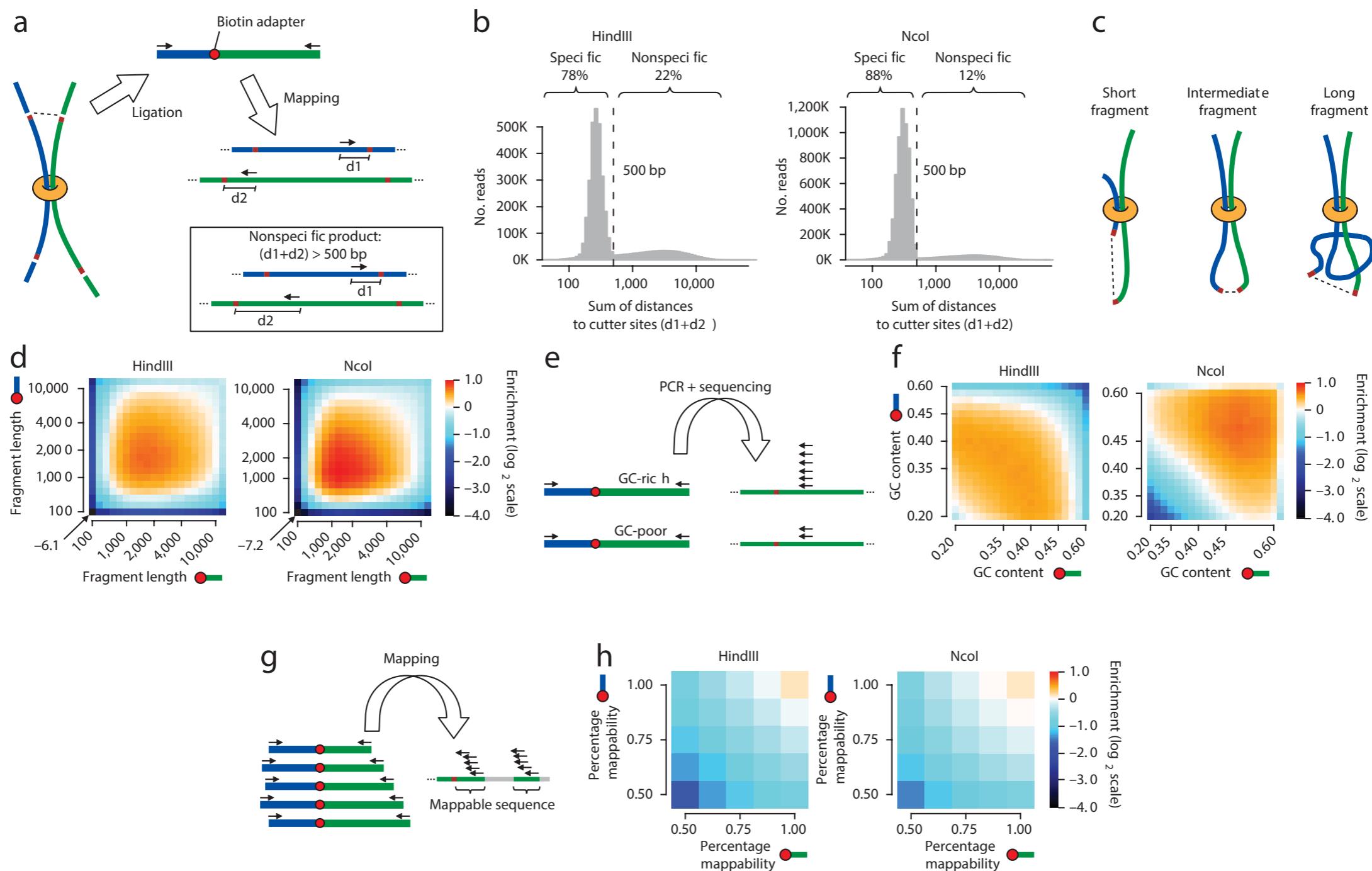
Giving a uniform “visibility” to all the rows and columns



Day 3

Normalizing HiC data (a la Tanay)

Yaffe, E., & Tanay, A. (2011). Nature Genetics, 43(11), 1059–1065

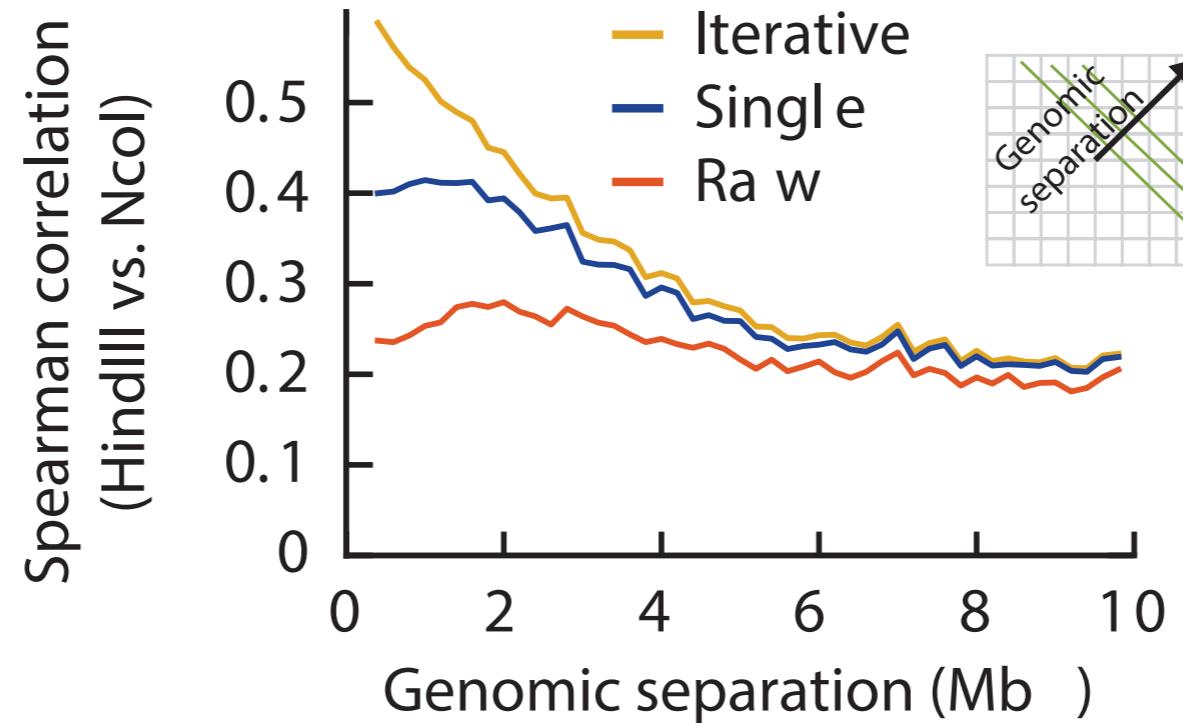
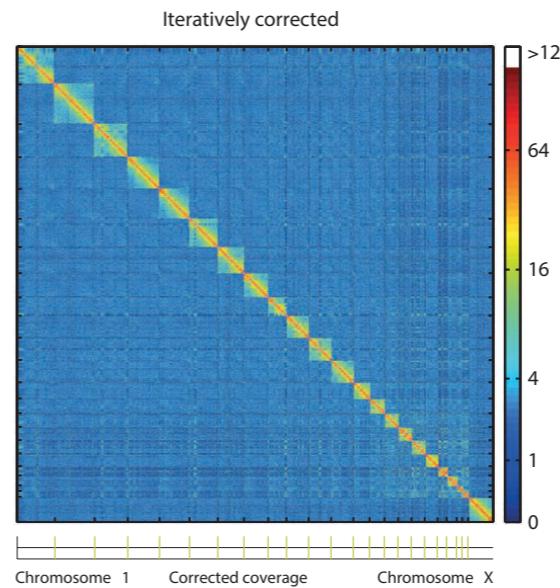


Normalizing HiC data (a la Mirny)

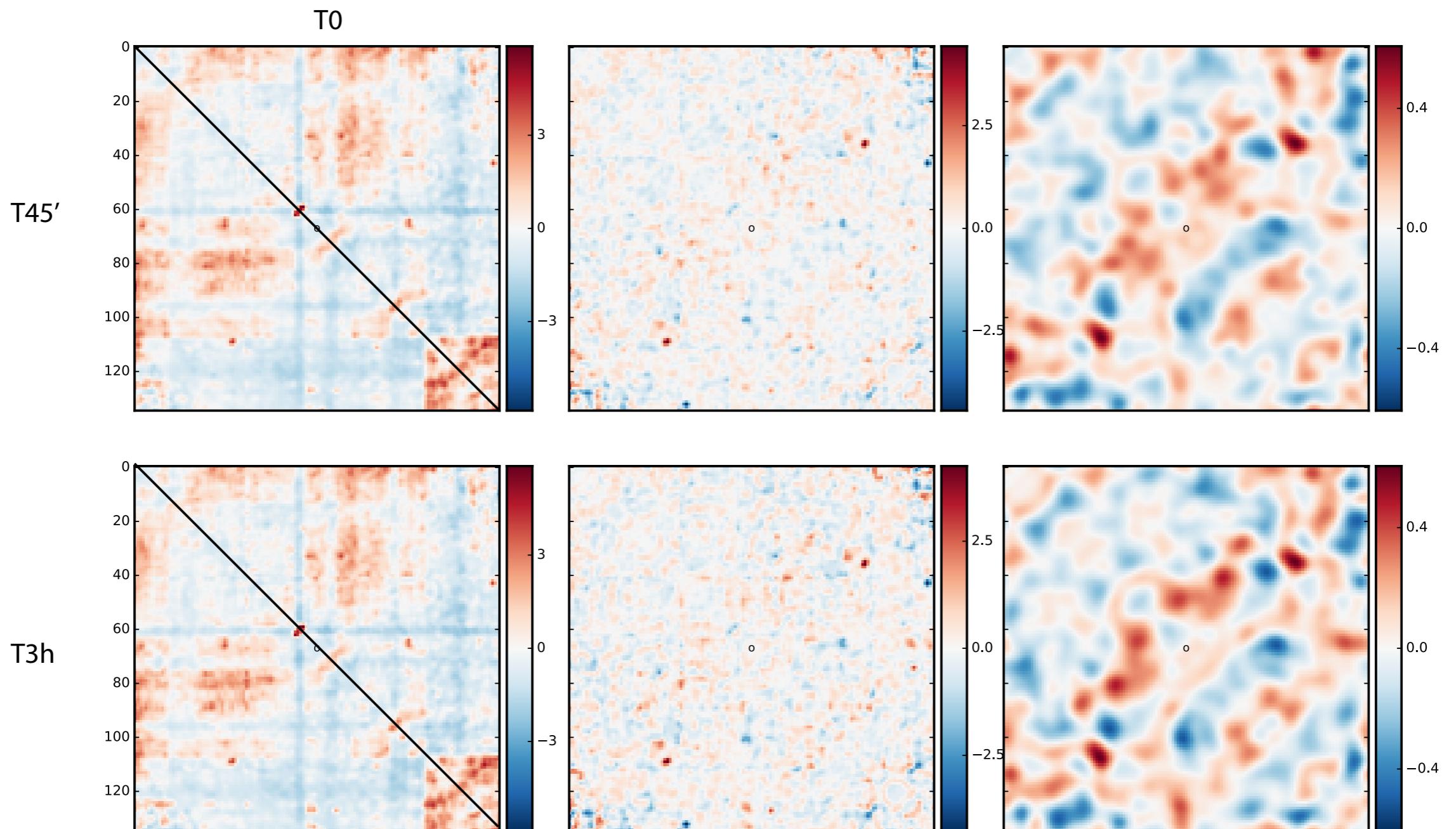
Imakaev, M., Fudenberg, G., McCord, R. P., Naumova, N., Goloborodko, A., Lajoie, B. R., et al. (2012). Nature Methods, 9(10), 999–1003.

$$O_{ij} = B_i B_j T_{ij}$$

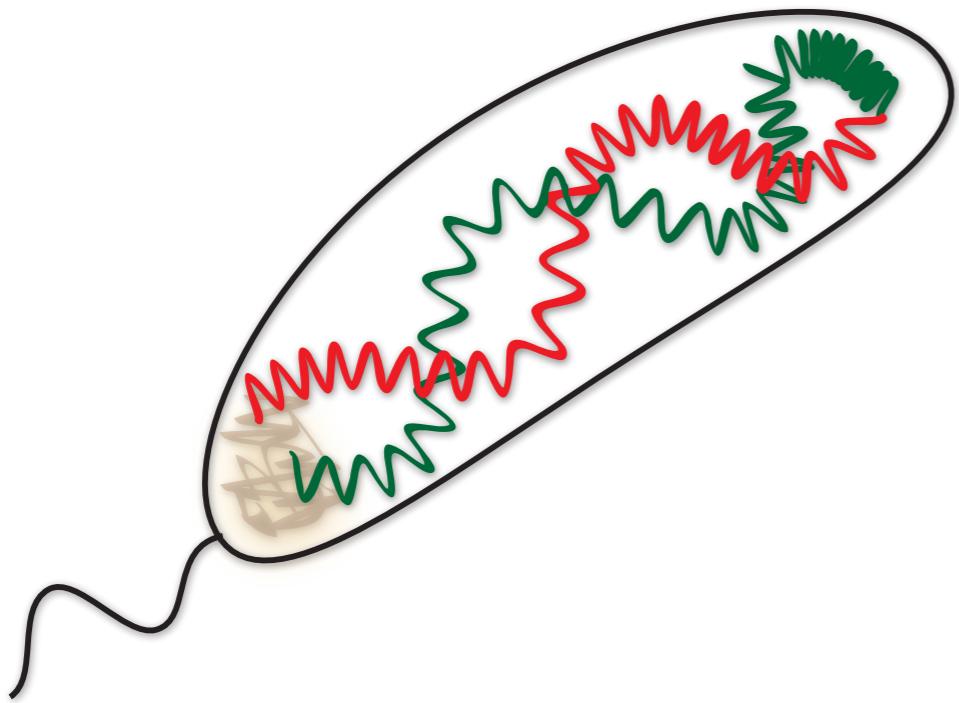
$$\sum_{i=1, |i-j|>1}^N T_{ij} = 1$$



Z-score differences (DekkerLab)



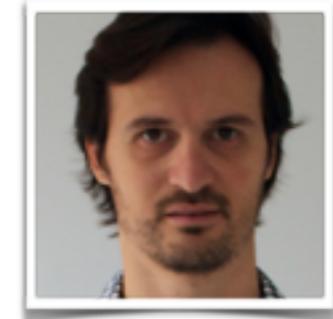
Caulobacter crescentus genome



Mark Umbarger
PhD fellow
Harvard



Esteban Toro
PhD fellow
Stanford

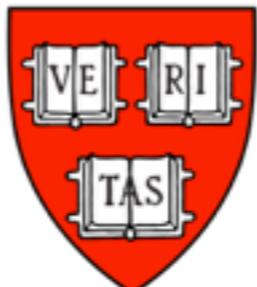


Davide Baù
Postdoctoral fellow
CIPF



Job Dekker

Program in Gene Function and Expression
Department of Biochemistry and Molecular Pharmacology
University of Massachusetts Medical School
Worcester, MA, USA



George M. Church

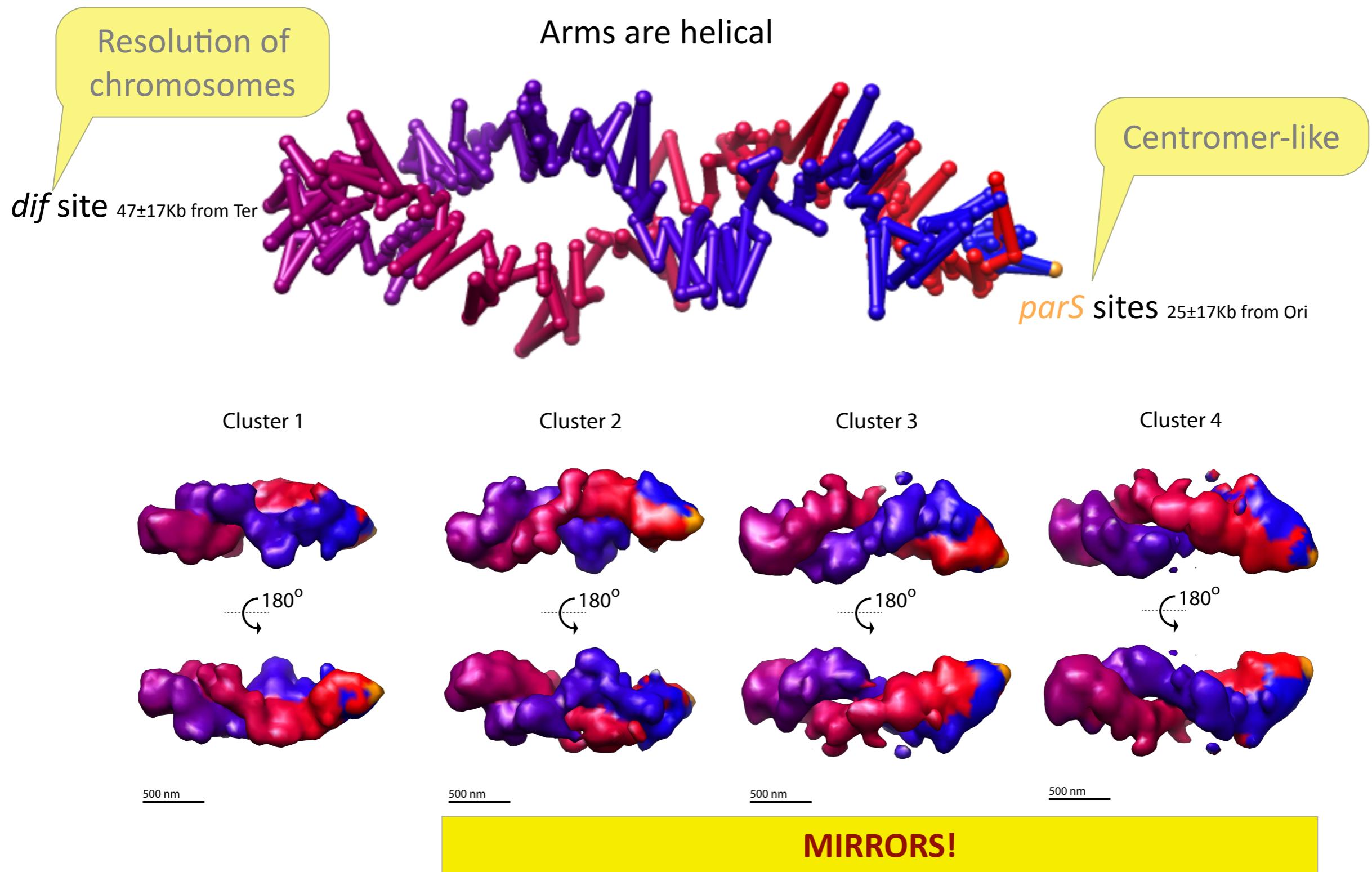
Department of Genetics,
Harvard Medical School,
Boston, MA. USA



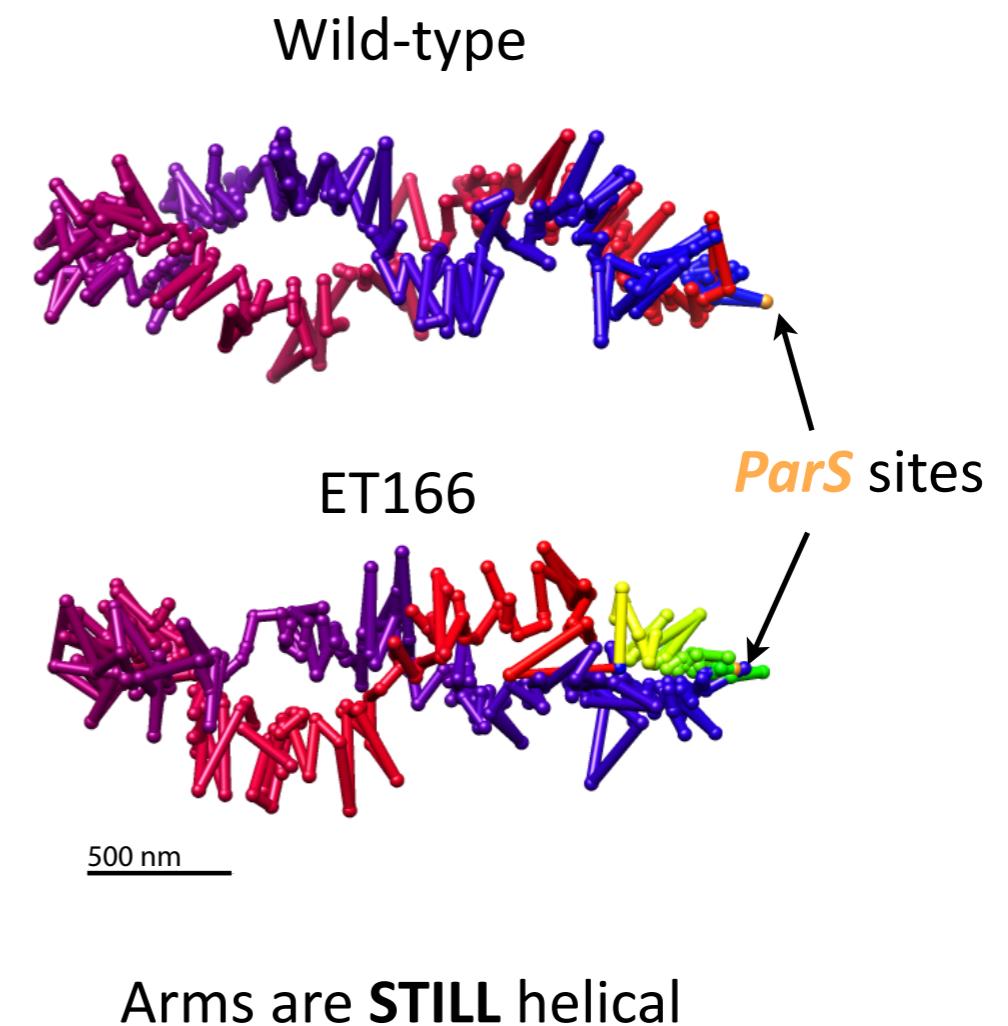
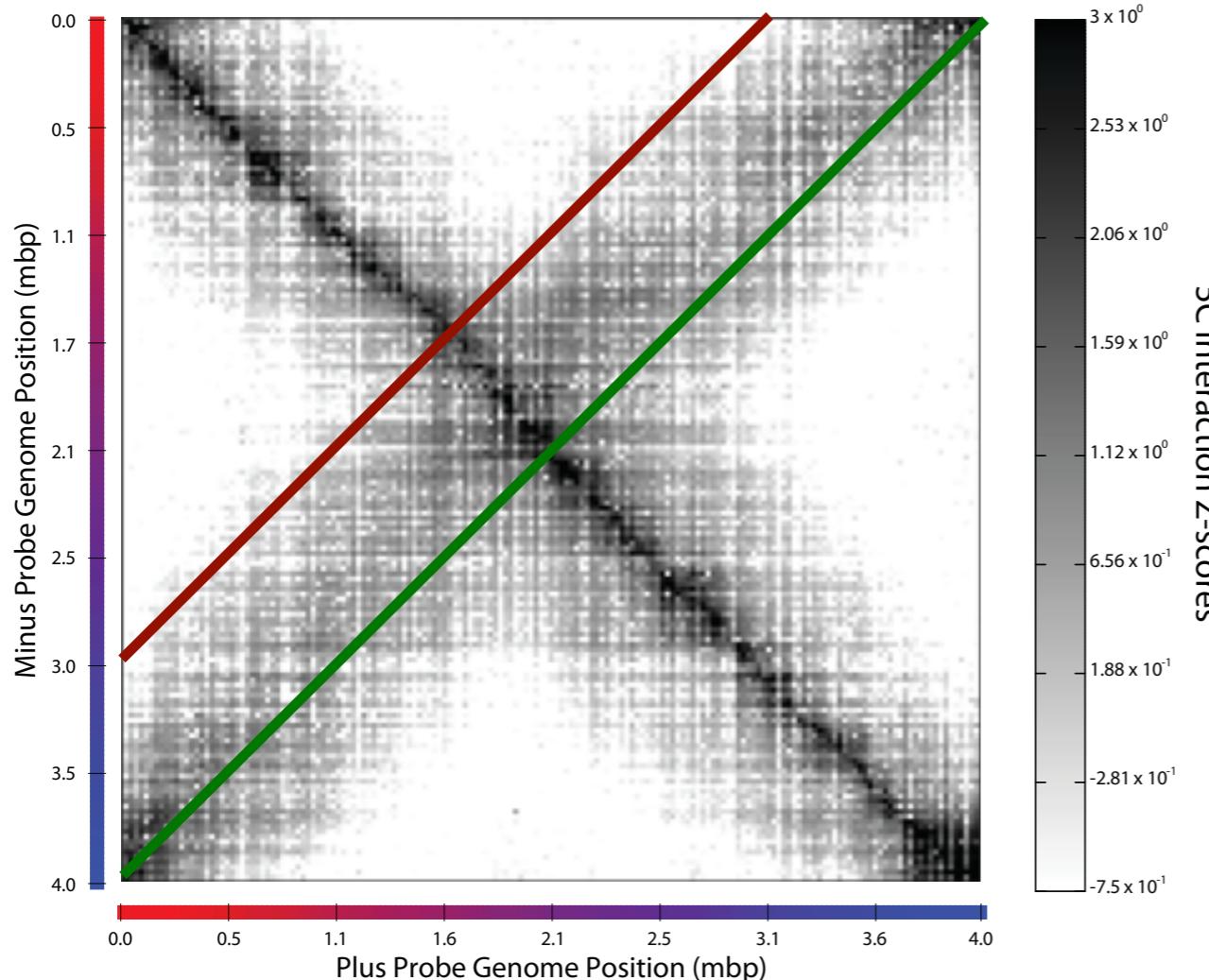
Lucy Shapiro

Department of Developmental Biology,
Stanford University School of Medicine,
Stanford, CA. USA

Genome organization in *Caulobacter crescentus*

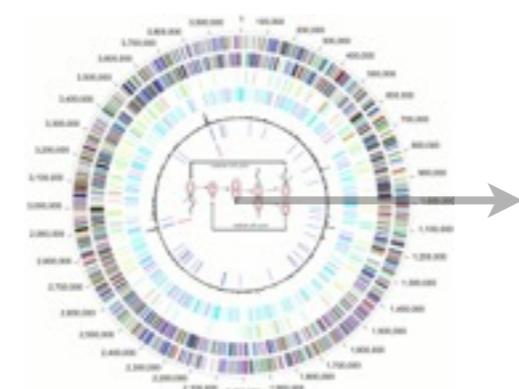
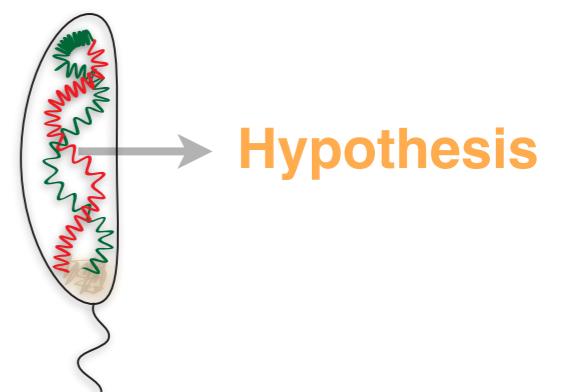
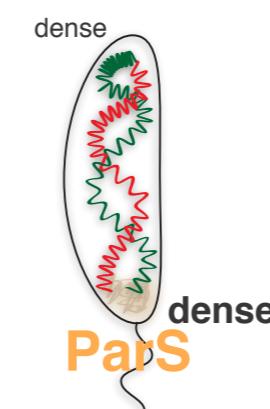
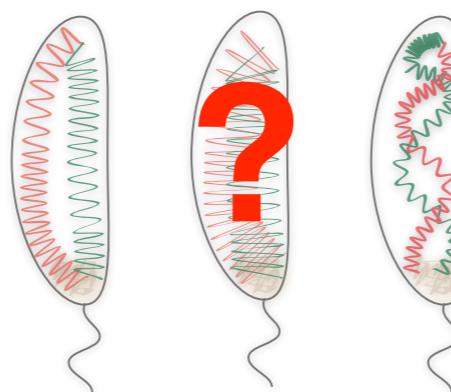
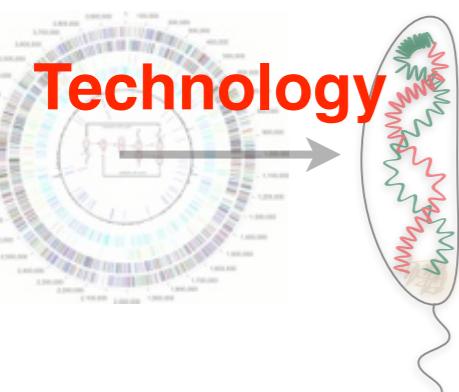
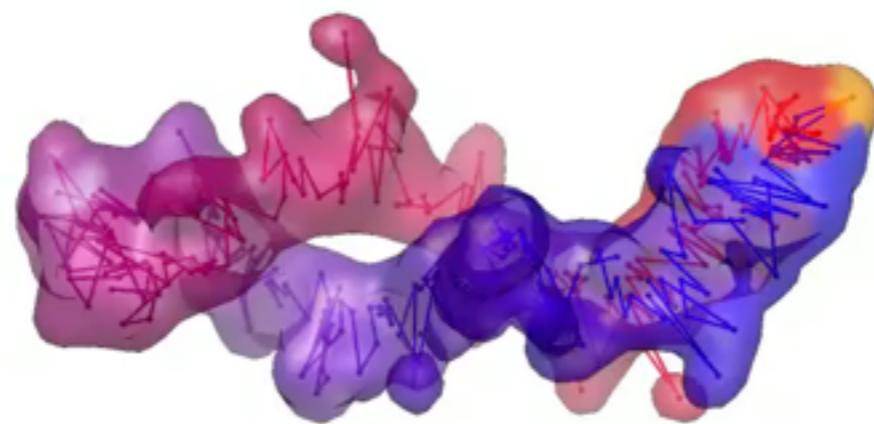


Moving the *parS* sites results in whole genome rotation!



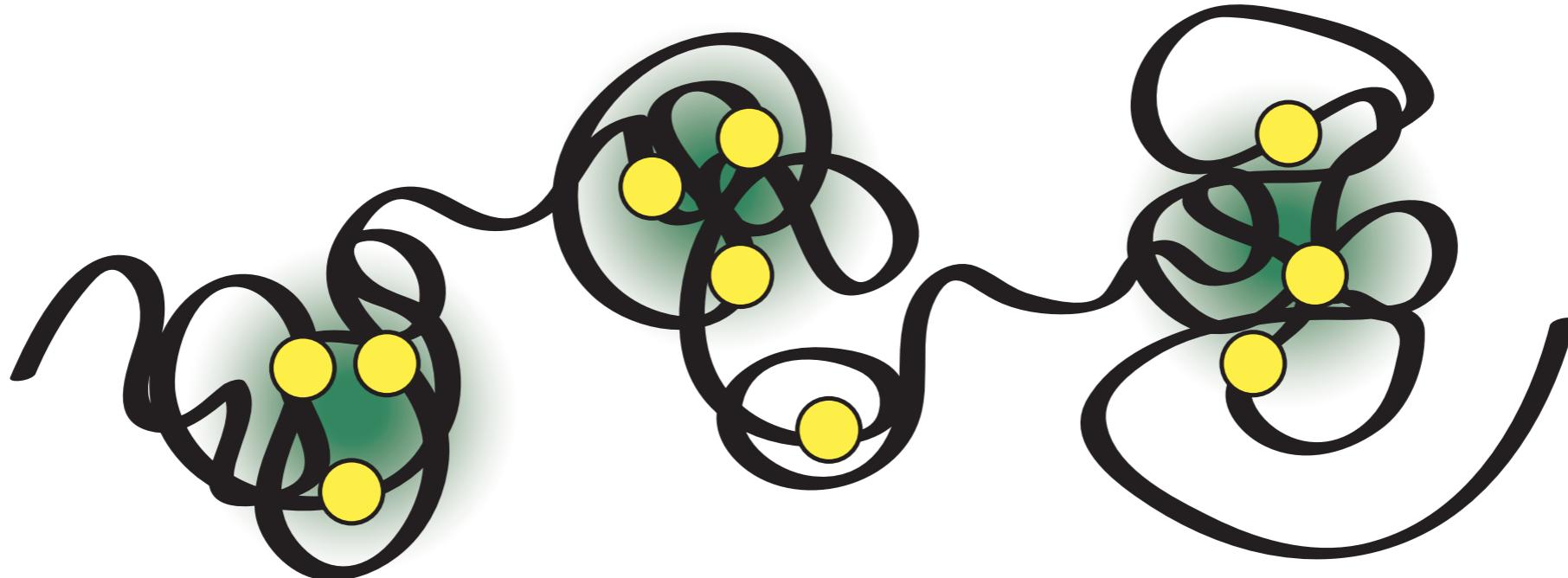
From Sequence to Function

Genome architecture in *Caulobacter*





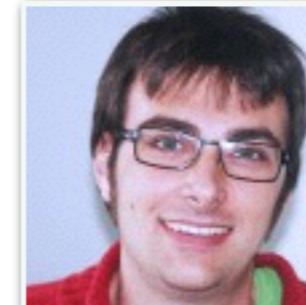
On TADs and hormones



François le Dily



Davide Baù



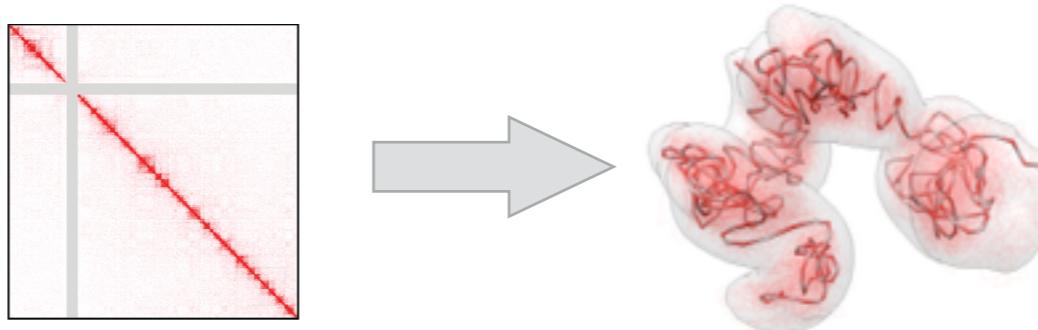
François Serra



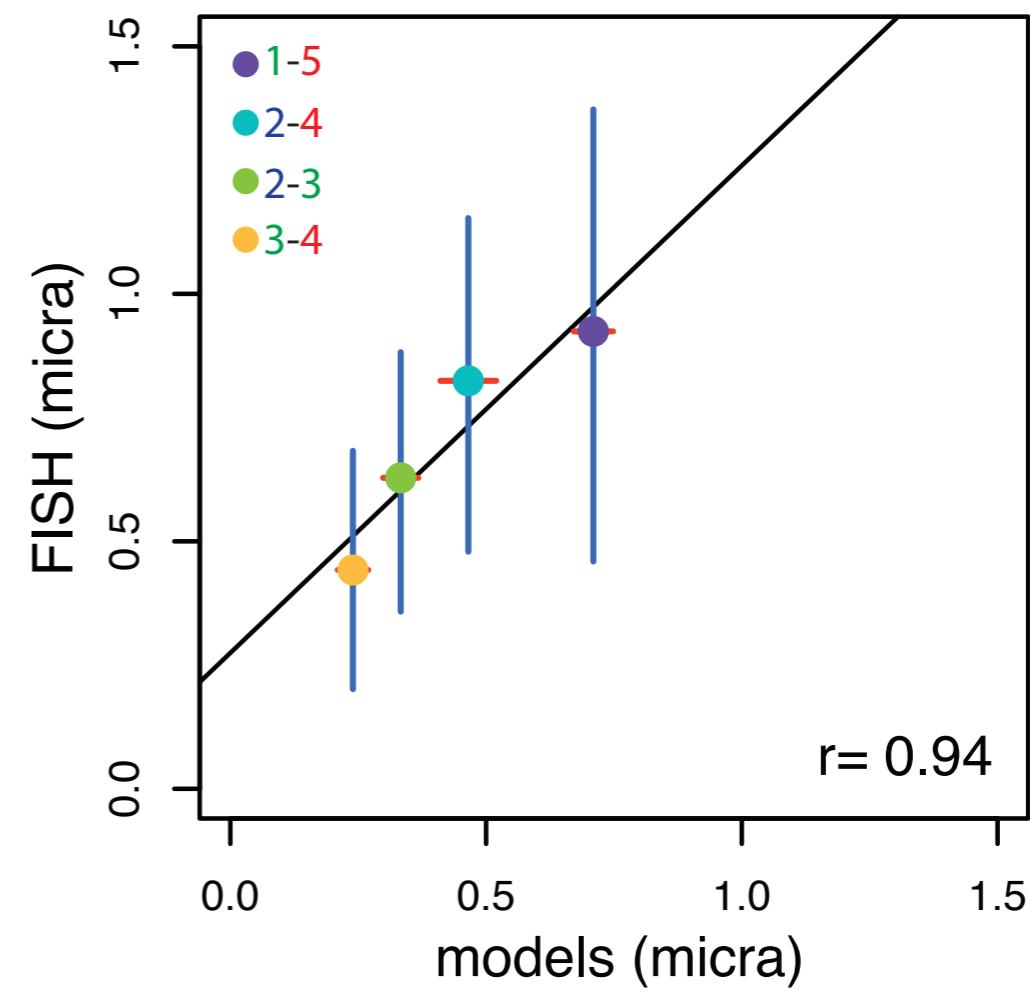
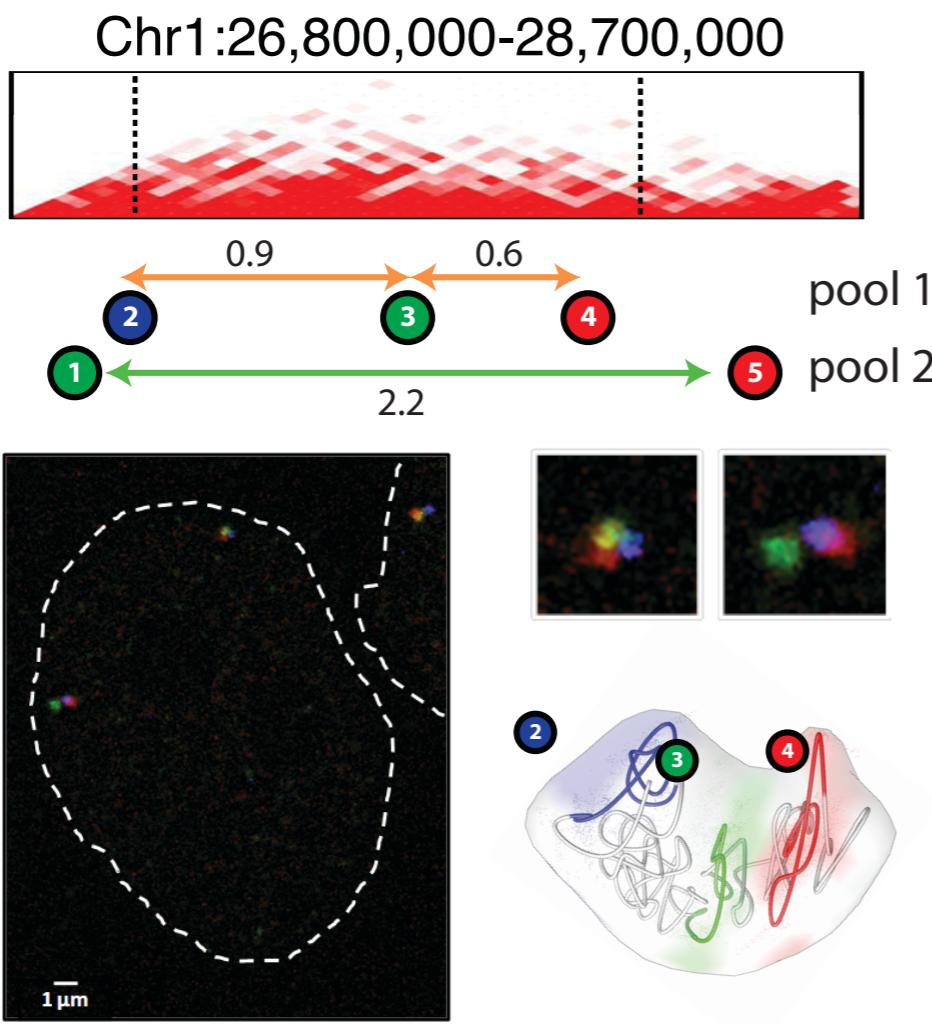
Miguel Beato & Guillaume Fillion

Gene Regulation, Stem Cells and Cancer
Centre de Regulació Genòmica
Barcelona, Spain

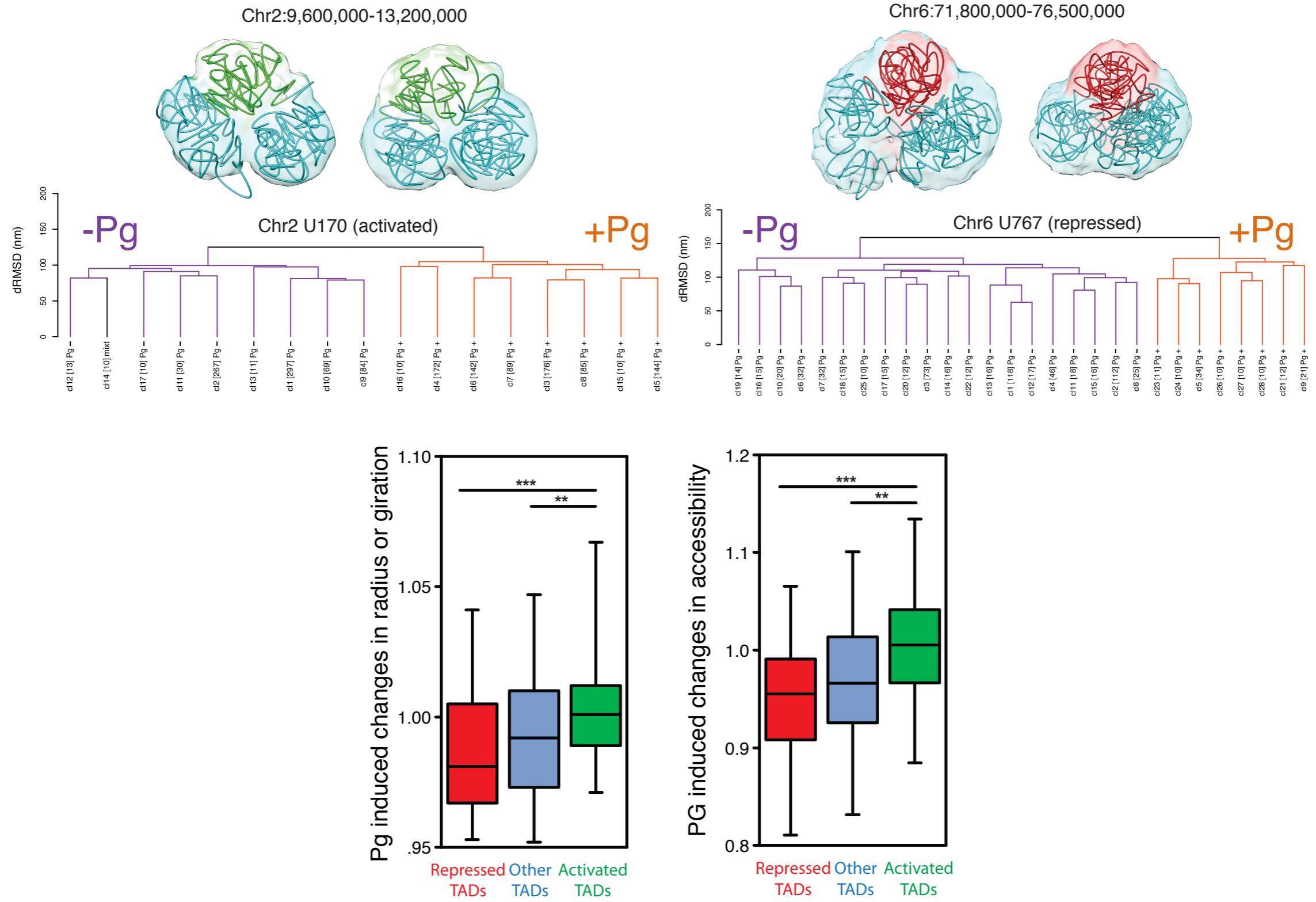
Modeling 3D TADs



61 genomic regions containing 209 TADs covering 267Mb



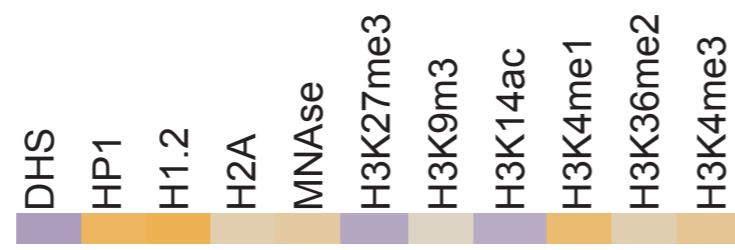
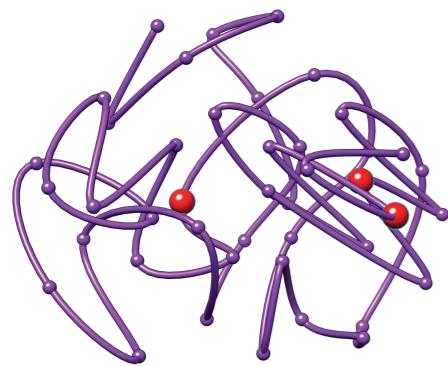
How TADs respond structurally to Pg?



Model for TAD regulation

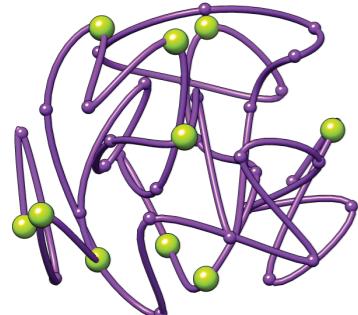
Repressed TAD

chr1 U41



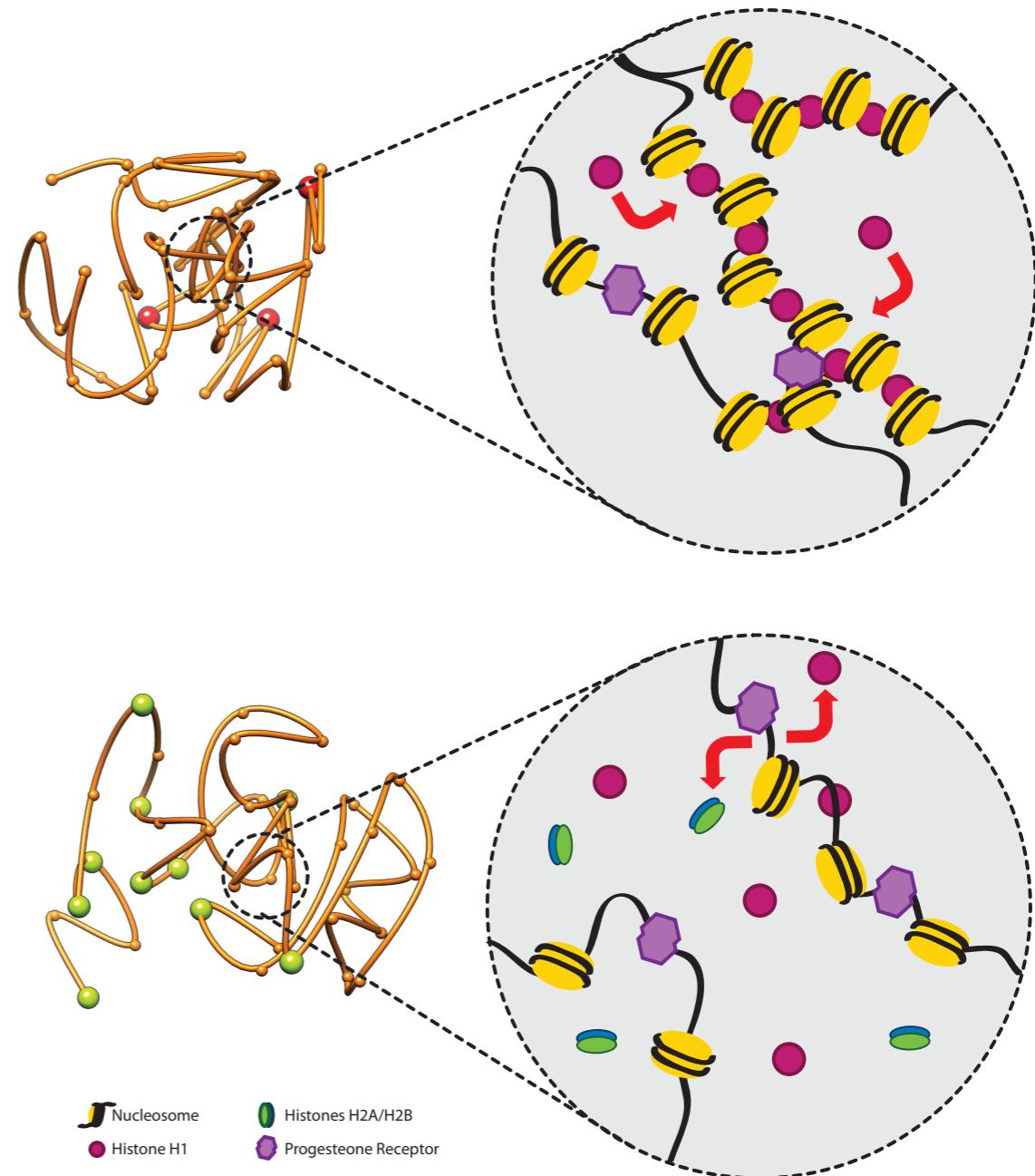
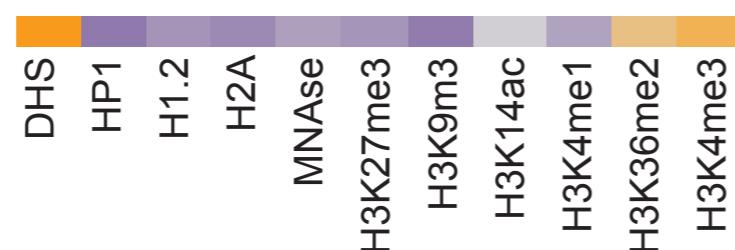
Activated TAD

chr2 U207



Structural transition

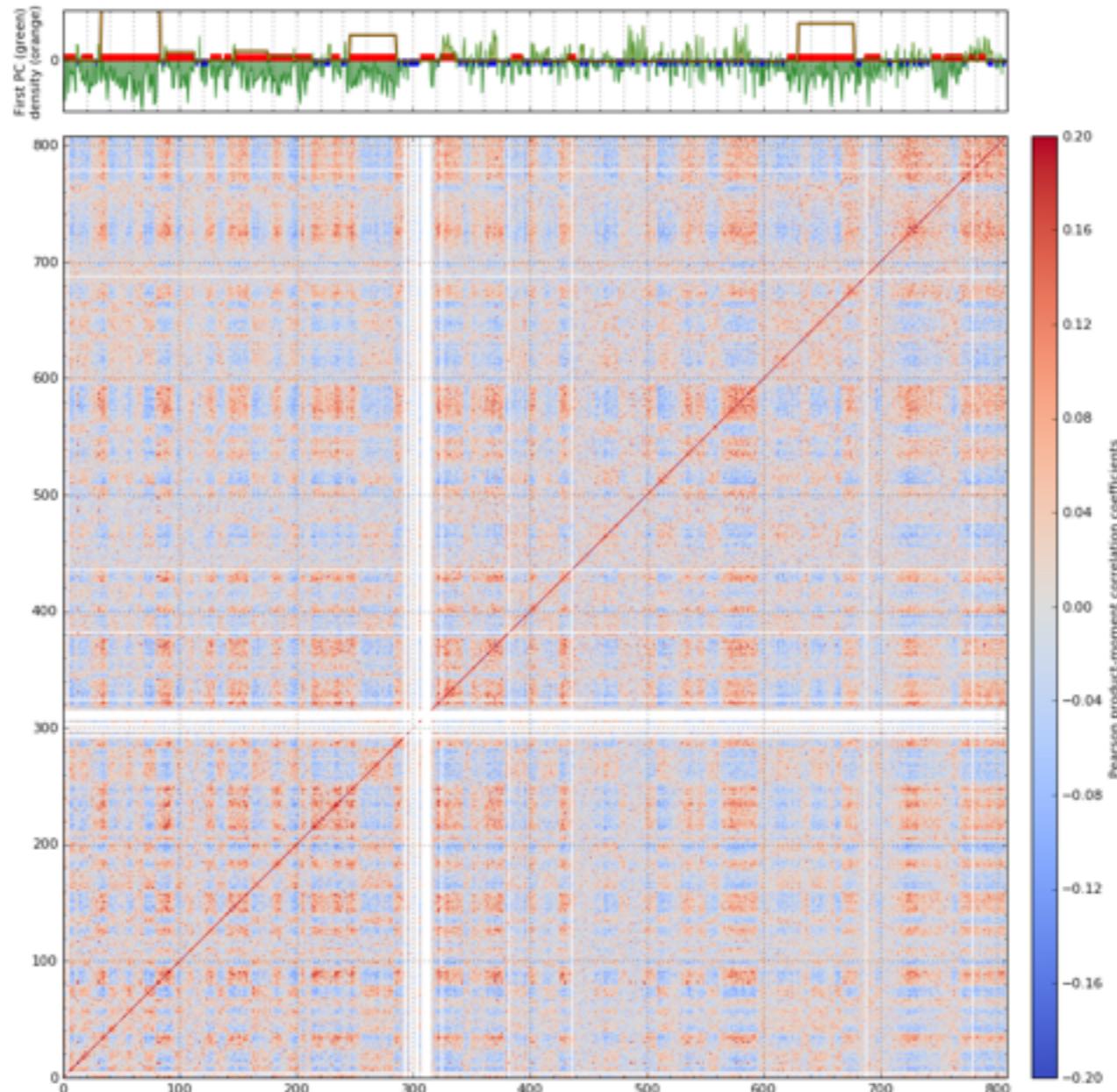
+Pg



A/B Compartment detection

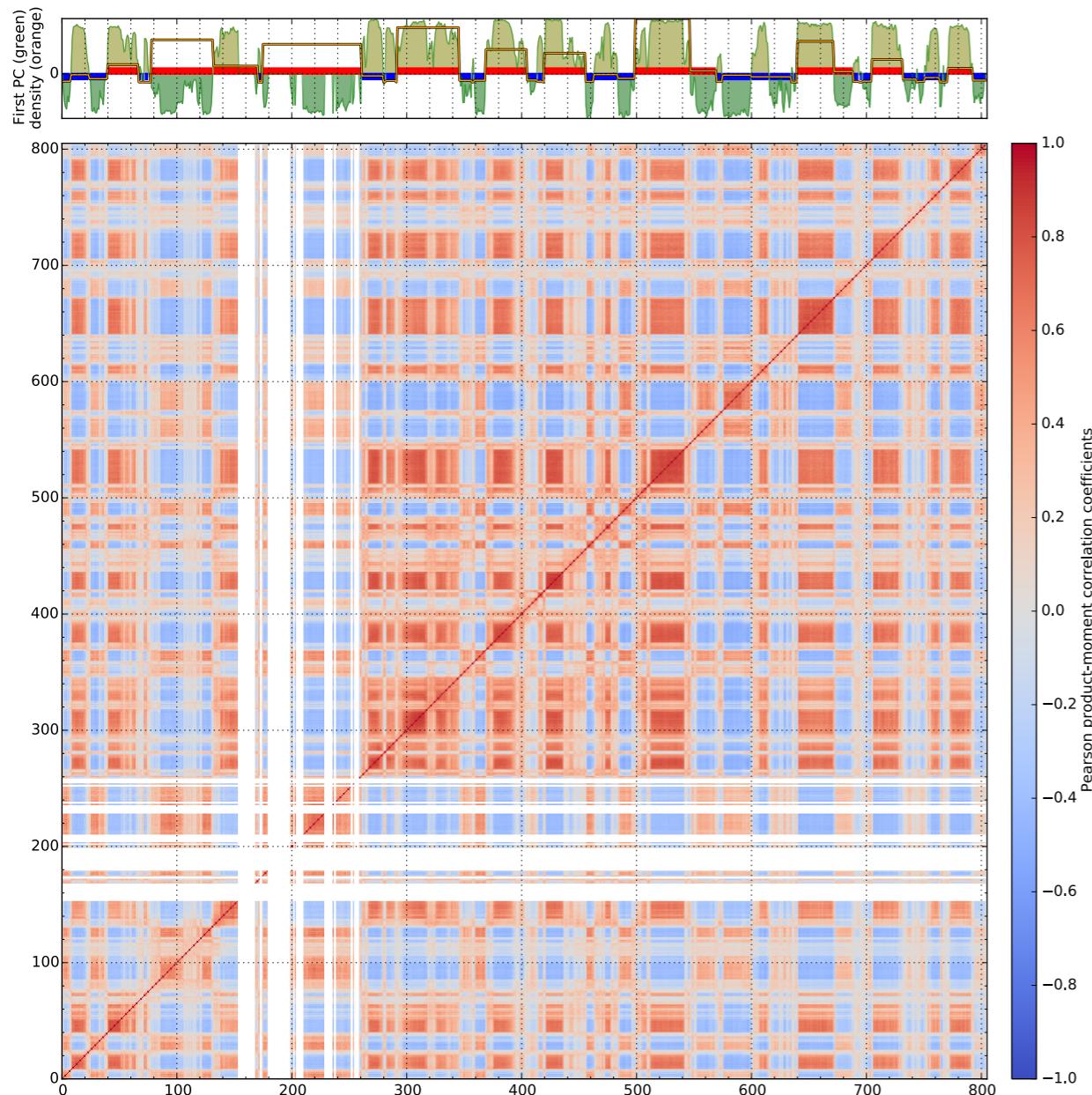
Chessboard pattern in the Hi-C matrix

```
In [33]: crm = '2'
corr = hic_data.find_compartments(
    show=True, crms=[crm], vmin=-0.2, vmax=0.2,
    savedata='results/fragment/both_{0}/05_segmenting/compartments_{1}_{2}.tsv'.format(rep, cr
m, reso))
```



A/B Compartment detection

Chessboard pattern in the Hi-C matrix



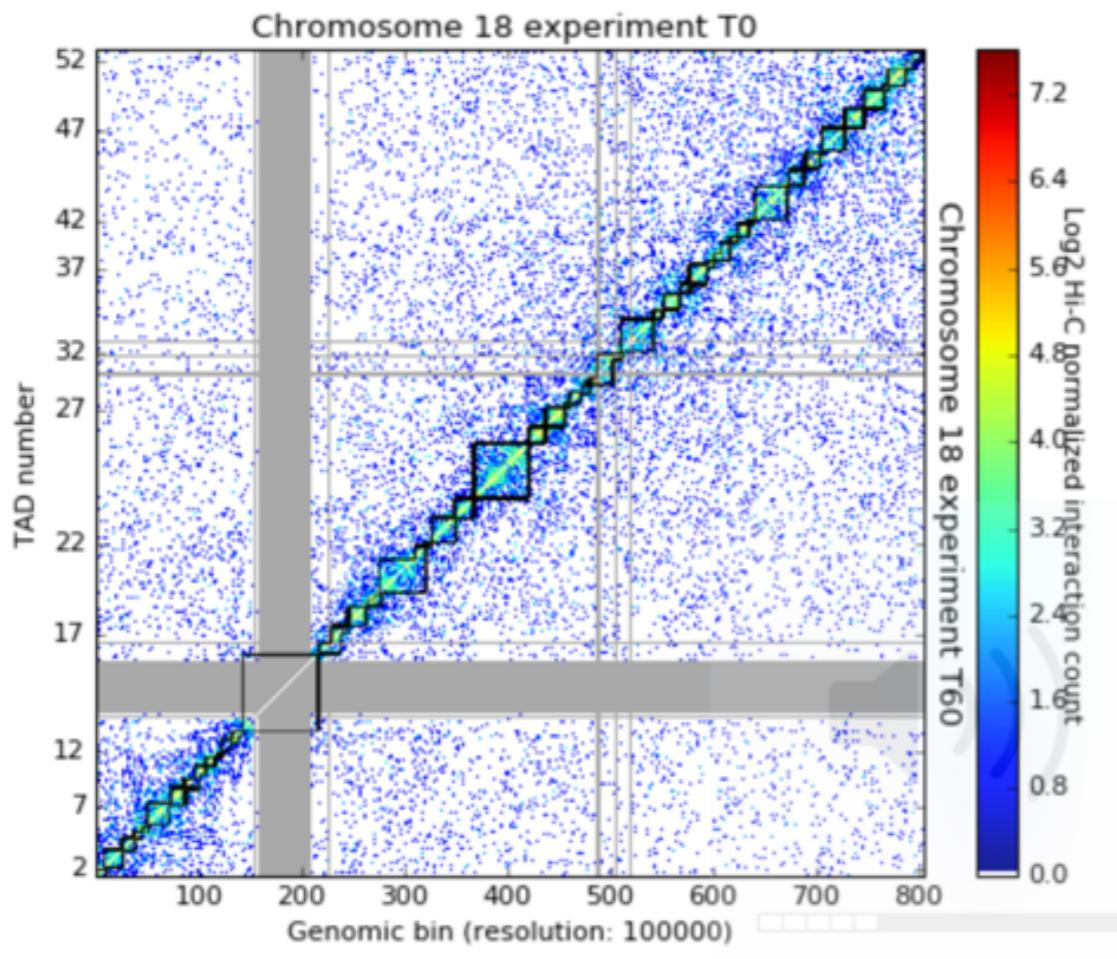
| # | start | end | density | type |
|----|-------|-----|---------|------|
| 18 | 1 | 8 | 0.42 | a |
| 18 | 9 | 10 | 0.02 | I |
| 18 | 11 | 20 | 1.06 | b |
| 18 | 21 | 24 | 0.11 | I |
| 18 | 25 | 26 | 0.02 | a |
| 18 | 27 | 33 | 0.32 | A |
| 18 | 34 | 40 | 0.33 | a |
| 18 | 41 | 41 | 0.01 | I |
| 18 | 42 | 52 | 1.12 | b |
| 18 | 53 | 70 | 2.94 | I |
| 18 | 71 | 72 | 0.06 | a |
| 18 | 73 | 76 | 0.10 | I |

TADs detection

Squares along the diagonal of the Hi-C matrix

```
In [50]: crm.find_tad(['T0', 'T60'], n_cpus=8)
```

```
In [54]: crm.visualize([('T0', 'T60')], normalized=True, paint_tads=True, )
```

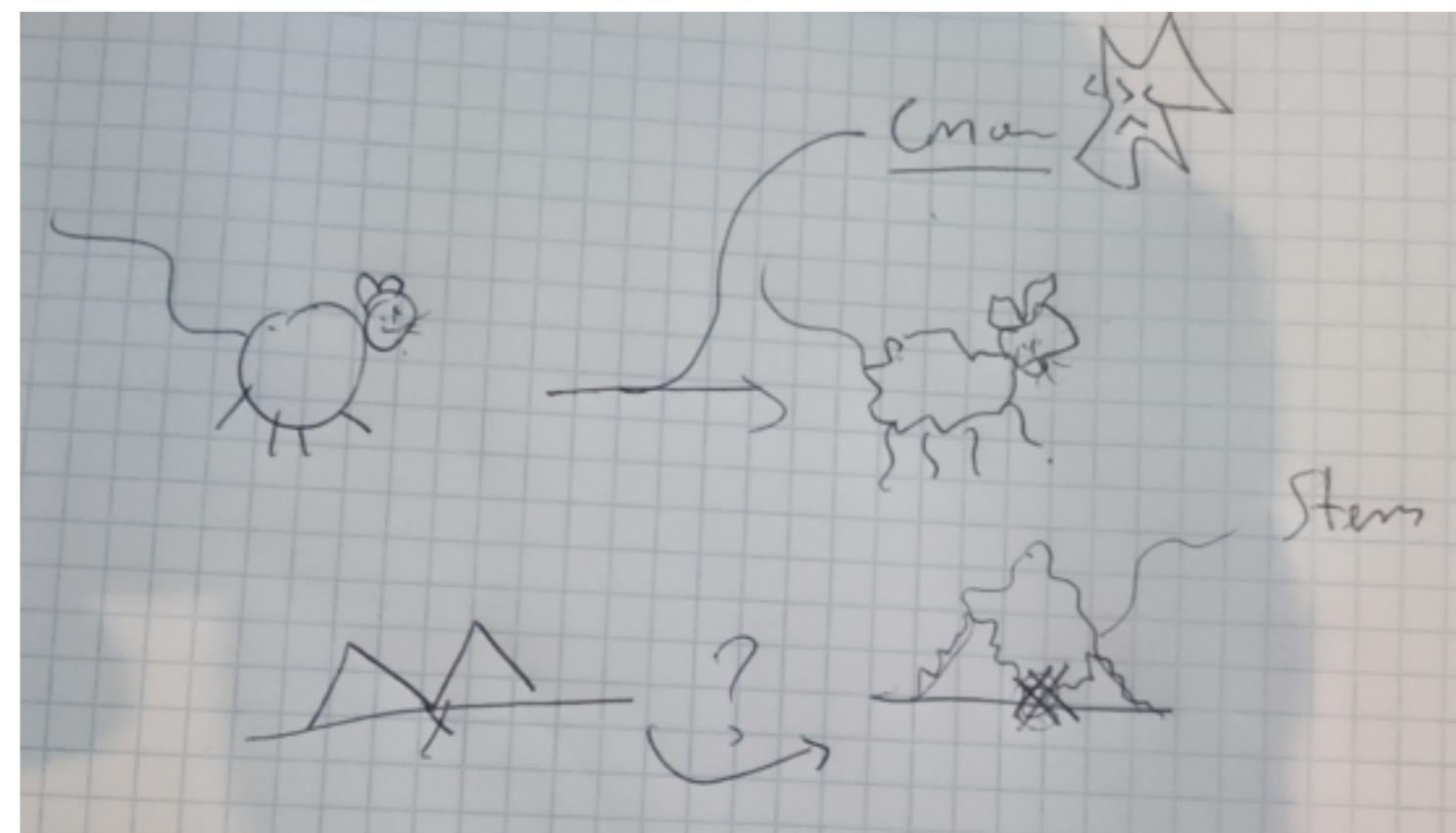
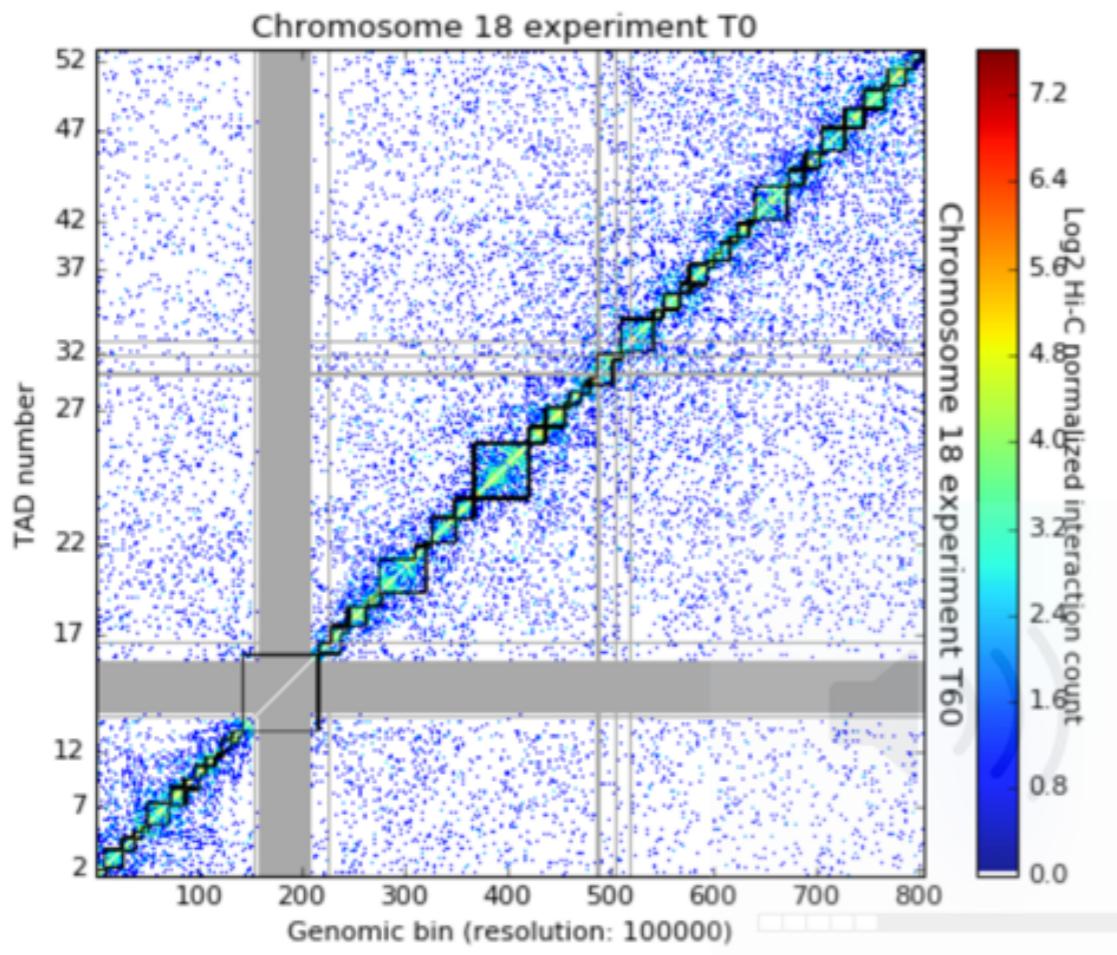


TADs detection

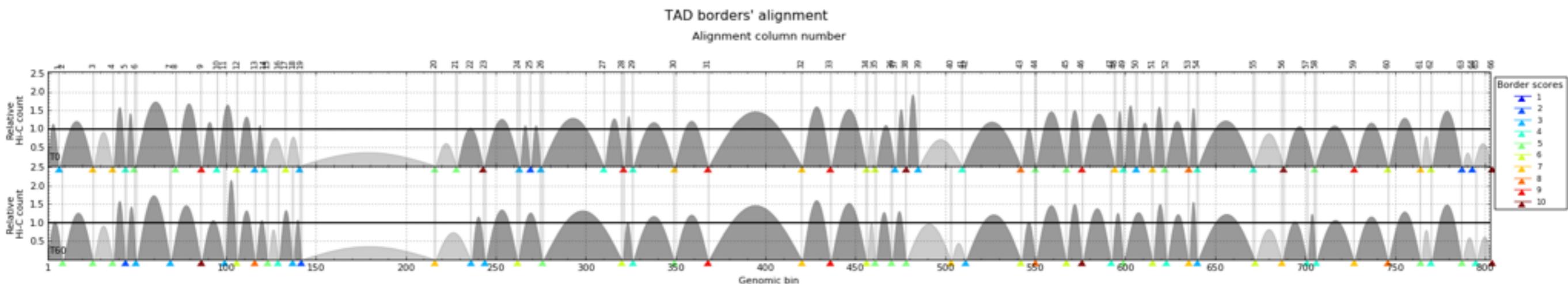
Squares along the diagonal of the Hi-C matrix

```
In [50]: crm.find_tad(['T0', 'T60'], n_cpus=8)
```

```
In [54]: crm.visualize([('T0', 'T60')], normalized=True, paint_tads=True, )
```



TADs analysis



```
In [11]: stats
```

```
Out[11]: (0.5151515151515151, 0.0, 0.8301886792452831, 0.8269230769230769)
```

```
In [13]: print 'Alignment score: %.3f, p-value: %.4f\n proportion of borders of T0 found in T60: %.3f, of T60 in T0 %.3f' % stats
```

```
Alignment score: 0.515, p-value: 0.0000
proportion of borders of T0 found in T60: 0.830, of T60 in T0 0.827
```

Questions on Day3

1 - How comfortable you feel to propose a validation strategy for your models?
...to the point to ask money for it to your boss? :-)

Questions on Day3

2 - How comfortable you feel to find compartments and TADs?

Questions on Day3

3 - How comfortable you feel in using TADbit?