

http://tinyurl.com/UPF-3DG-Tutorial

PRE TUTORIAL

1. DOWNLOAD DATA TO YOUR HOME DIR

Go to <http://tinyurl.com/UPF-3DG-Tutorial>

2. UNTAR DATA IN YOUR HOME DIR

```
>tar -xvf tutorial_3DG.tar.gz
```

3. FROM YOUR HOME DIR! run:

```
>tutorial_3DG/00_install_tadbit.sh
```

TUTORIAL

1. ENTER THE TUTORIAL DIR

```
>cd tutorial_3DG
```

2. QUALITY PLOTS (seconds each command)

```
>tadbit map -w results --fastq FASTQs/Control_rep4.1_10M.fastq --read 1 --index db/dm6.gem --renz DpnII --skip_mapping
>tadbit map -w results --fastq FASTQs/Control_rep4.2_10M.fastq --read 2 --index db/dm6.gem --renz DpnII --skip_mapping
```

2. MAP (~8 minutes each command, ~15 minutes all)

```
>tadbit map -w results --fastq FASTQs/Control_rep4.1_10M.fastq --read 1 --index db/dm6.gem --renz DpnII
>tadbit map -w results --fastq FASTQs/Control_rep4.2_10M.fastq --read 2 --index db/dm6.gem --renz DpnII
```

3. PARSE (~5 minutes)

```
>tadbit parse -w results --compress_input --genome db/dm6.fa
```

4. FILTER (~4 minutes)

```
>tadbit filter -w results
```

4. NORMALIZE @100Kb (~2 minutes)

```
>tadbit normalize -w results -r 100000 --min_count 100
```

5. SEGMENT THE GENOME @100Kb A/B + TADs (~2 minutes)

```
>tadbit segment -w results -r 100000 -c 2L 2R 3L 3R 4 X
```

6. MODEL (~20 minutes)

Re-normalize @10kb

```
>tadbit normalize -w results -r 10000 --min_count 100
```

Get matrix @10kb

```
>tadbit bin -w results --resolution 10000 --coord 'X:100000000-110000000' --norm raw norm --plot --triangular
```

Find optimal parameters for modeling

```
>tadbit model -w results --reso 10000 --crm X --beg 100000000 --end 110000000 --project Tutorial --species 'Drosophila melanogaster' --assembly 'dm6' --cell kc167 --maxdist 200:200:100 --upfreq=-0.0:0.3:0.3 --lowfreq=-0.3:0.0:0.3 --dcutoff 2:4:1 --scale 0.01 --nmodels 10 --nkeep 10 --cpu 10 --optimize
```

Model with optimal parameters

```
>tadbit model -w results --reso 10000 --crm X --beg 100000000 --end 110000000 --project Tutorial --species 'Drosophila melanogaster' --assembly 'dm6' --cell kc167 --maxdist 200 --upfreq 0.3 --lowfreq 0 --dcutoff 2 --scale 0.01 --nmodels 50 --nkeep 25 --cpu 10 --model
```

Analyze models

```
>tadbit model -w results --analyze
```

Visualize models using the JSON file...

<http://sgt.cnag.cat/3dg/tadkit/demo.h/>

Structure determination of genomes and genomic domains by satisfaction of spatial restraints

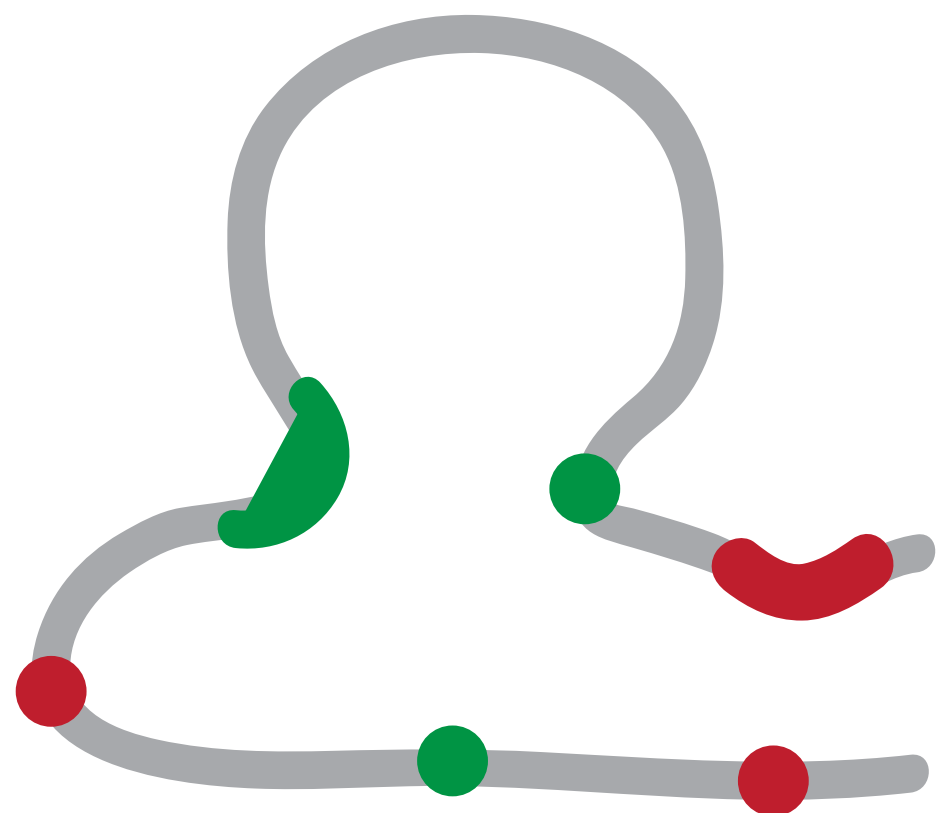
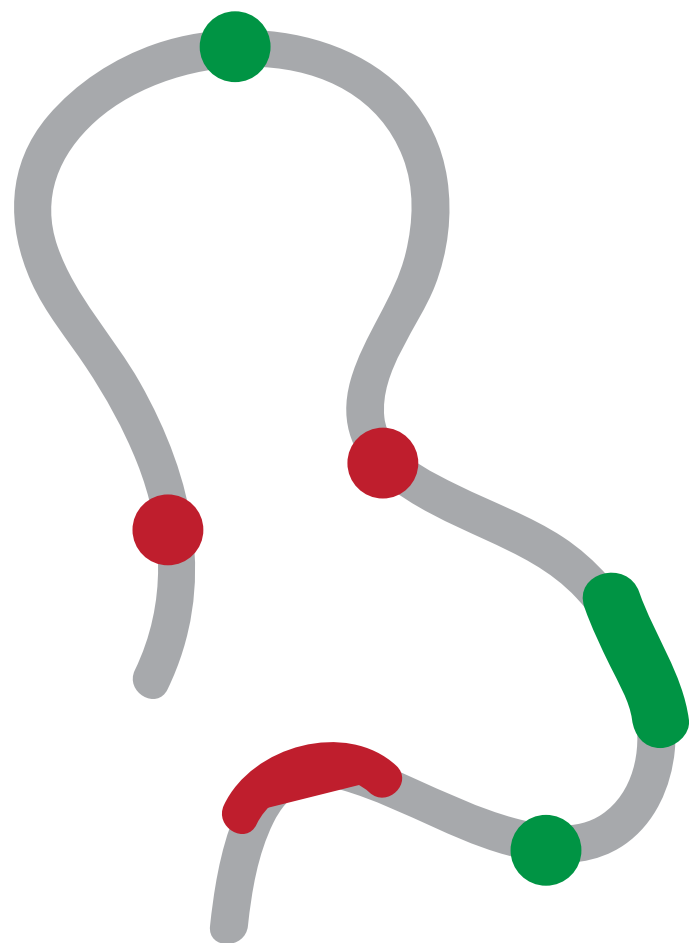
Marc A. Marti-Renom

Structural Genomics Group (ICREA, CNAG-CRG)

<http://marciuslab.org>
<http://3DGenomes.org>
<http://cnag.crg.eu>

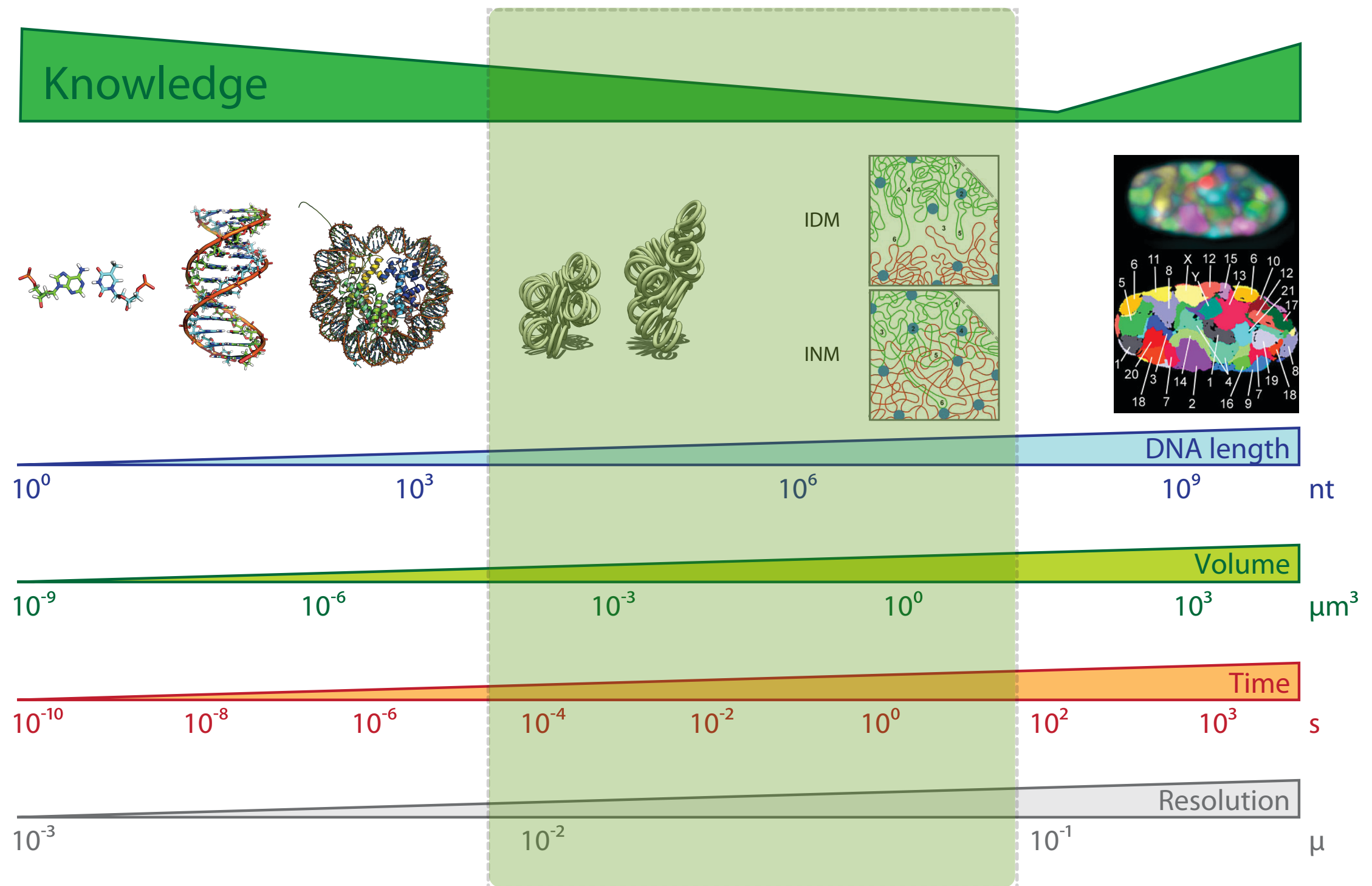
cnag **CRG**^R  **ICREA**





Resolution Gap

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

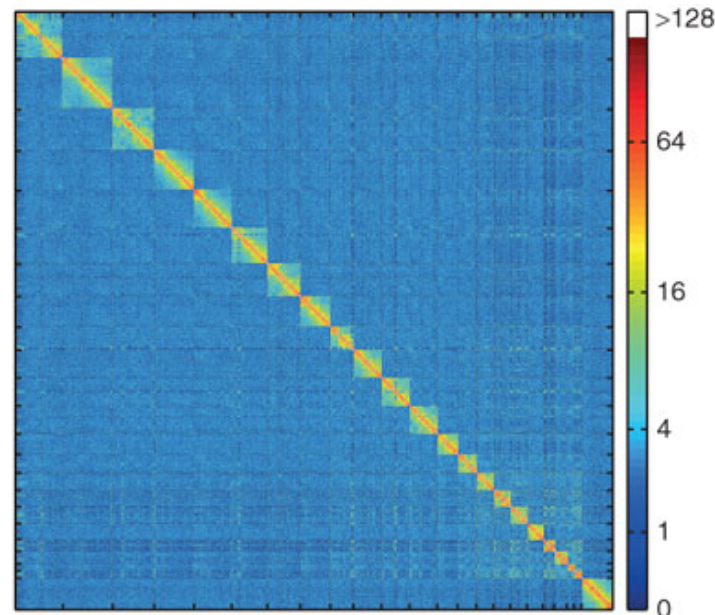
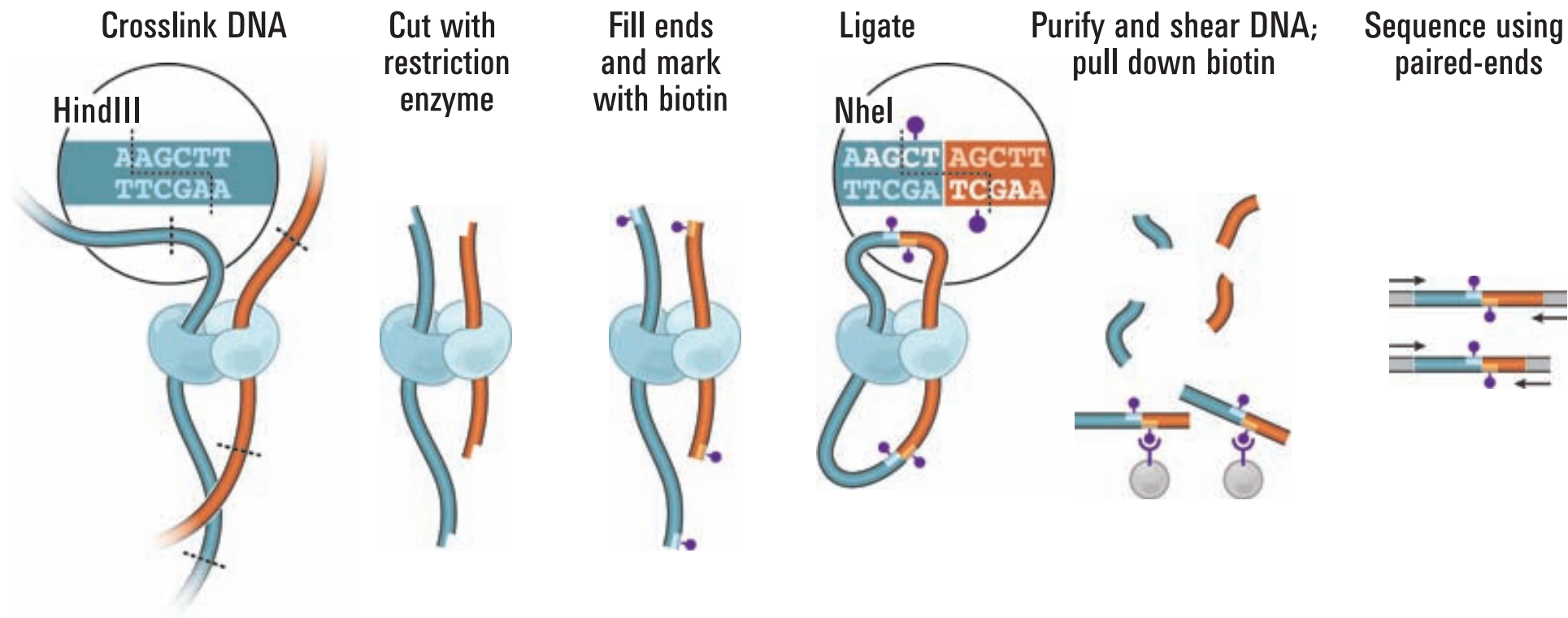




Got FASTQ?

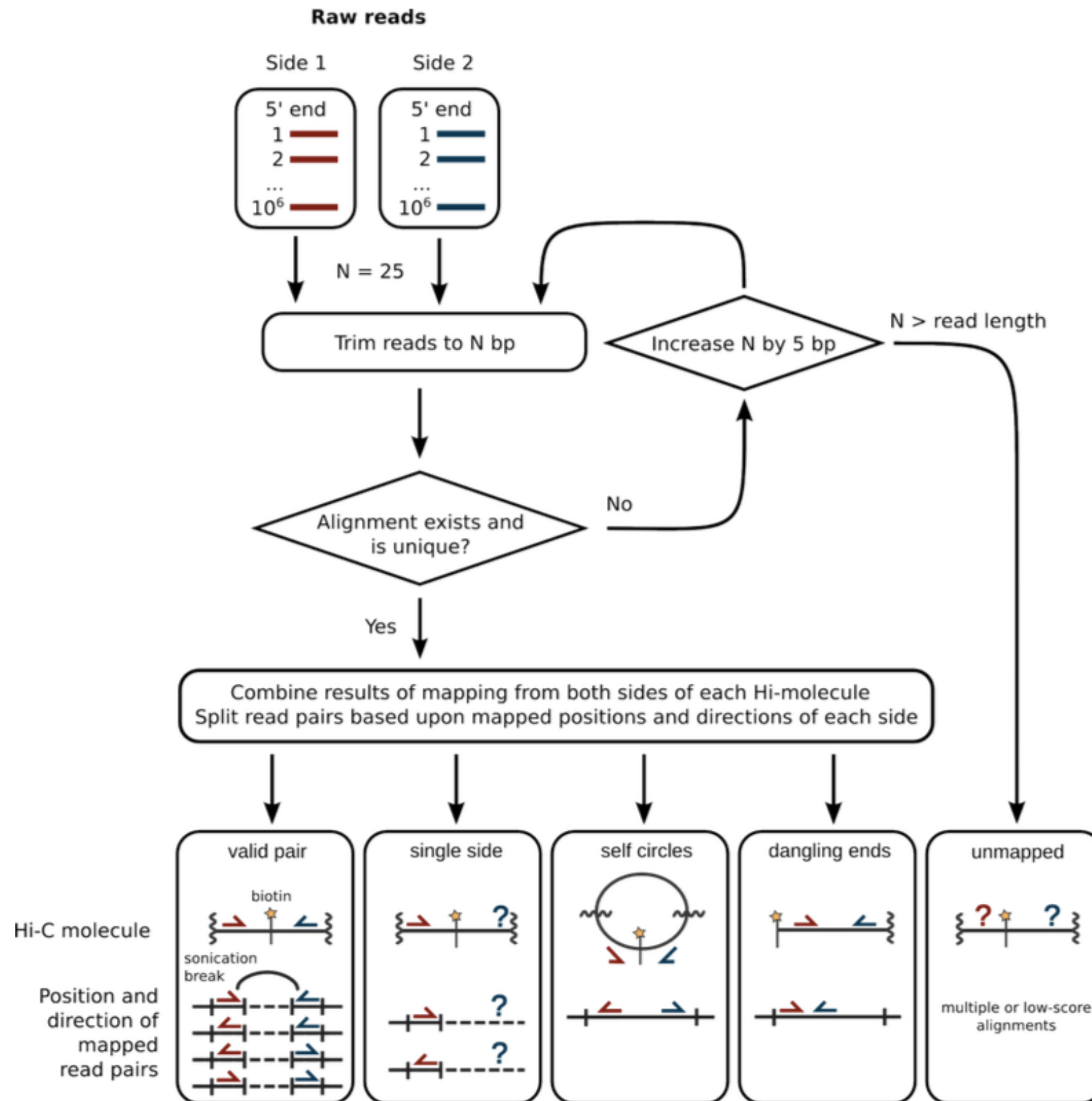
Hi-C experiment

Lieberman-Aiden, E., et al. (2009). *Science*, 326(5950), 289–293.



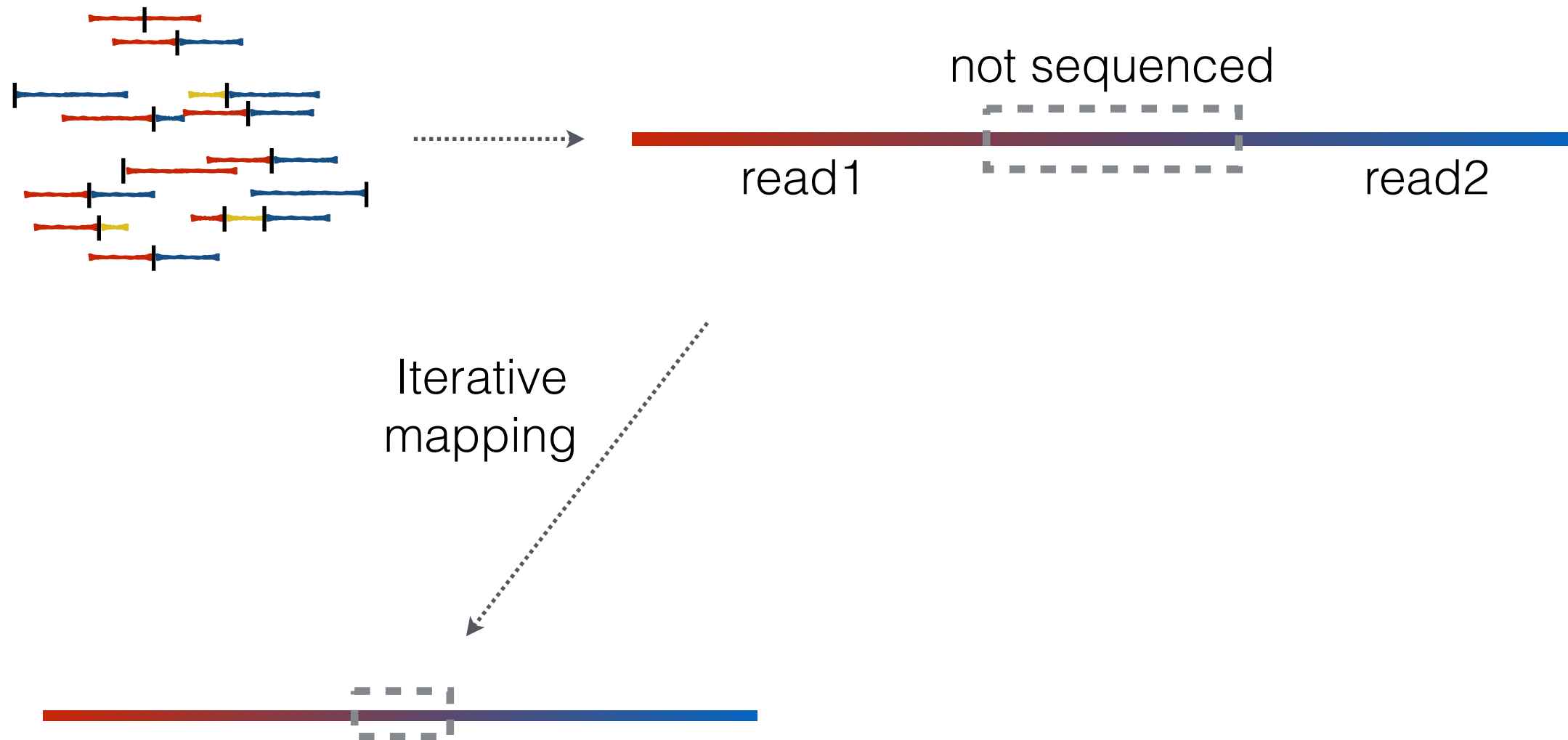
Mapping & Filtering

Imakaev, M. V et al. (2012). Nature Methods, 9(10), 999–1003.



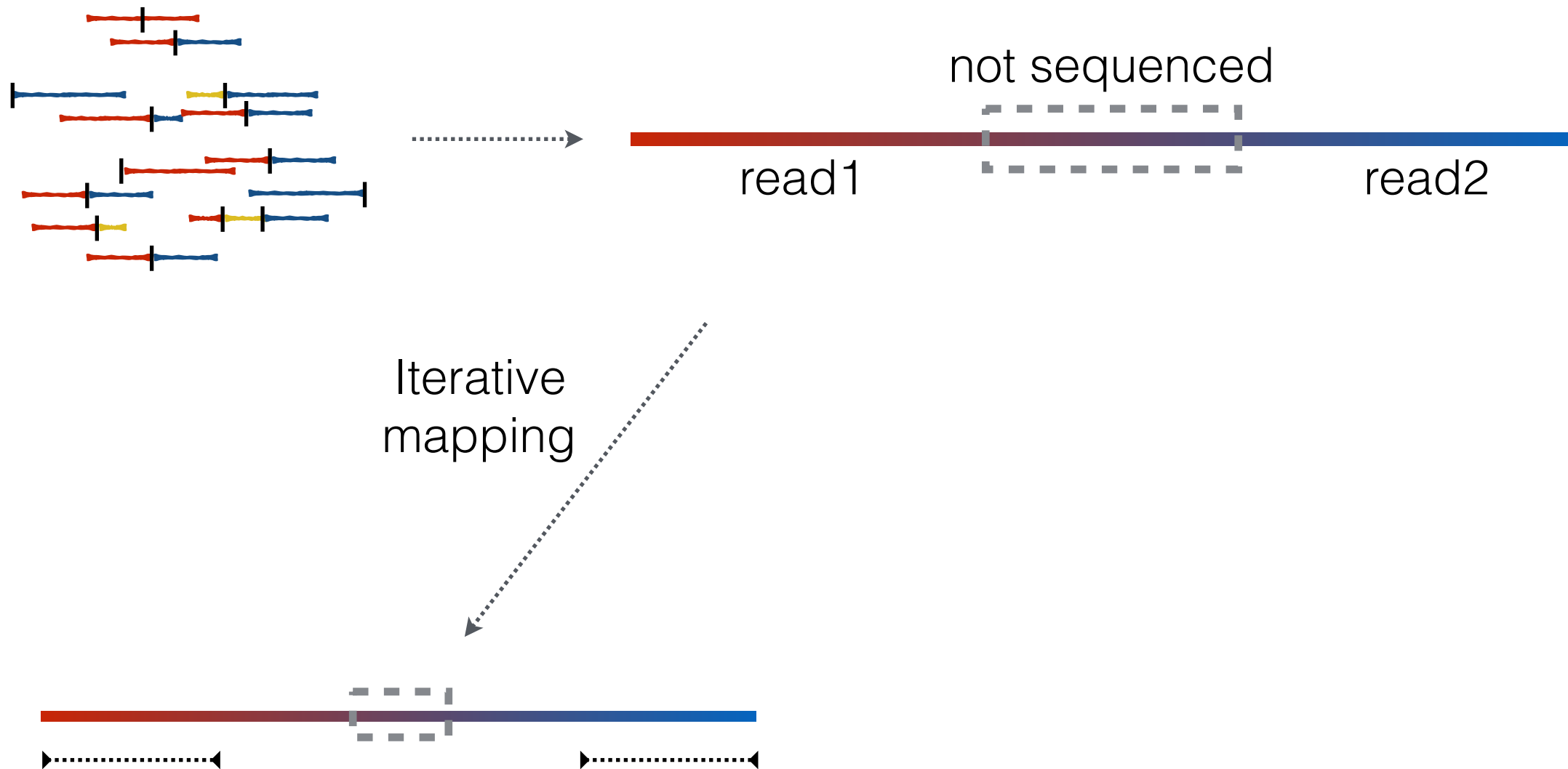
Mapping @TADbit

Serra, Baù, et al. (2017). PLOS CompBio



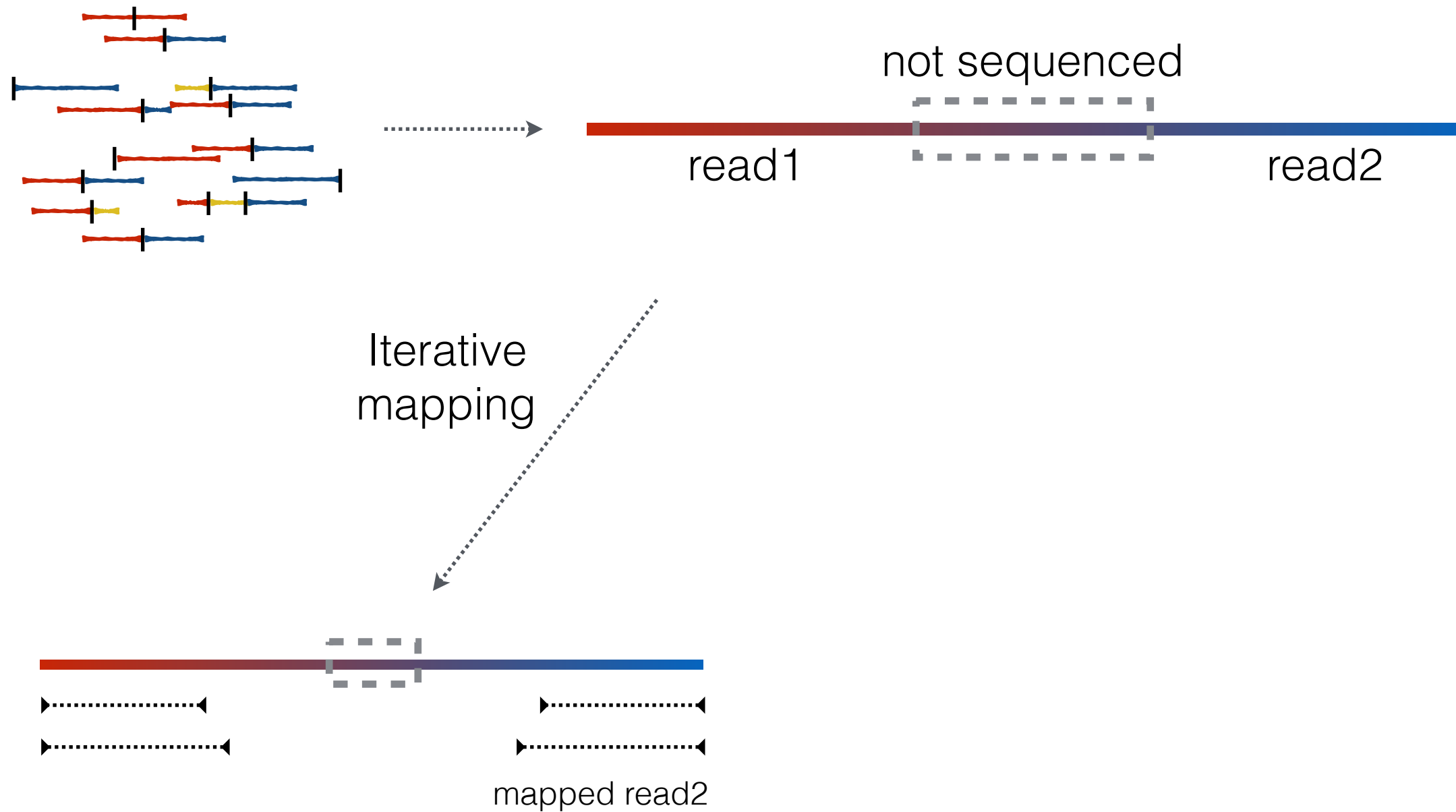
Mapping @TADbit

Serra et al. (2017). PLOS CompBio



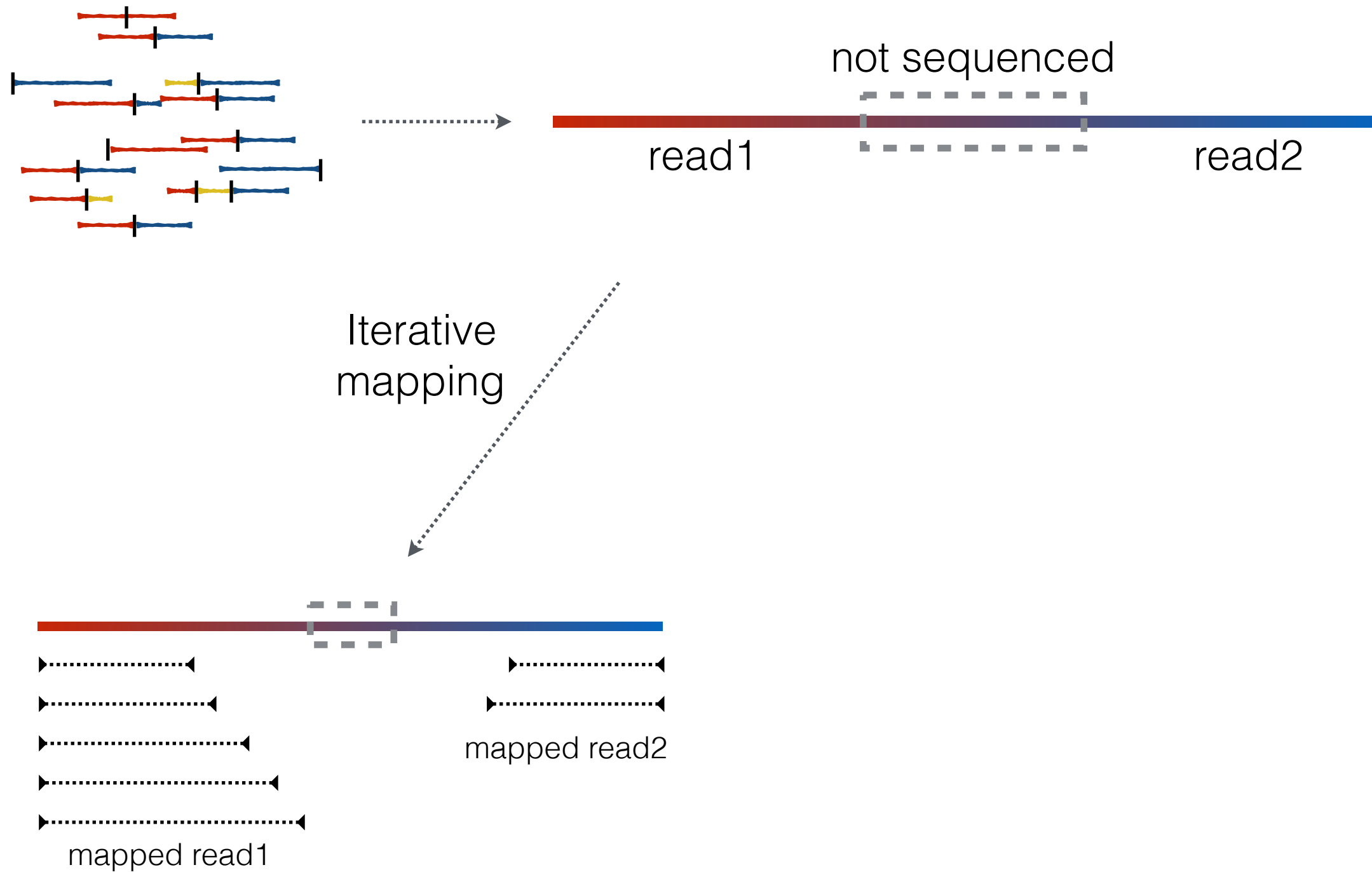
Mapping @TADbit

Serra et al. (2017). PLOS CompBio



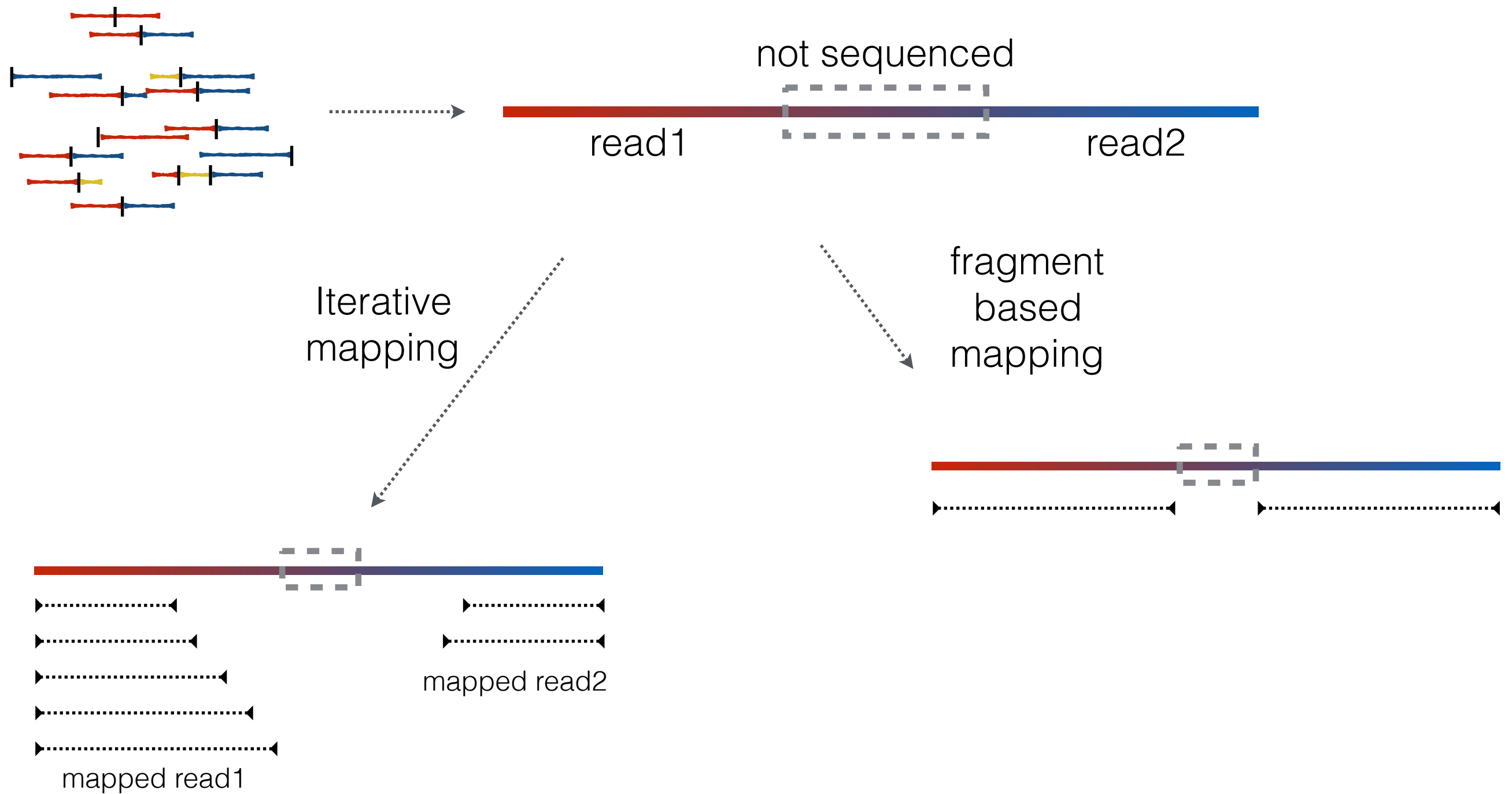
Mapping @TADbit

Serra et al. (2017). PLOS CompBio



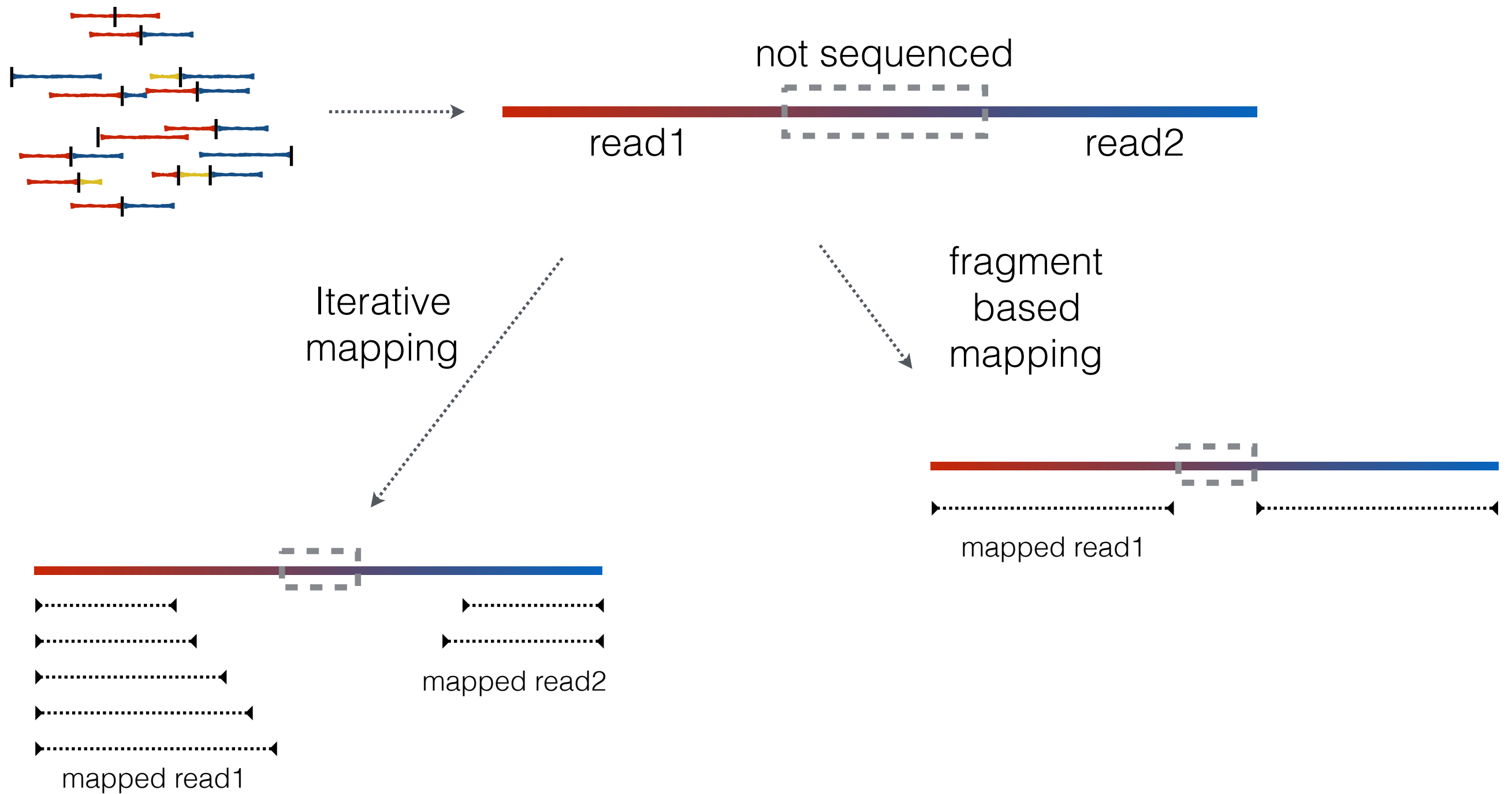
Mapping @TADbit

Serra et al. (2017). PLOS CompBio



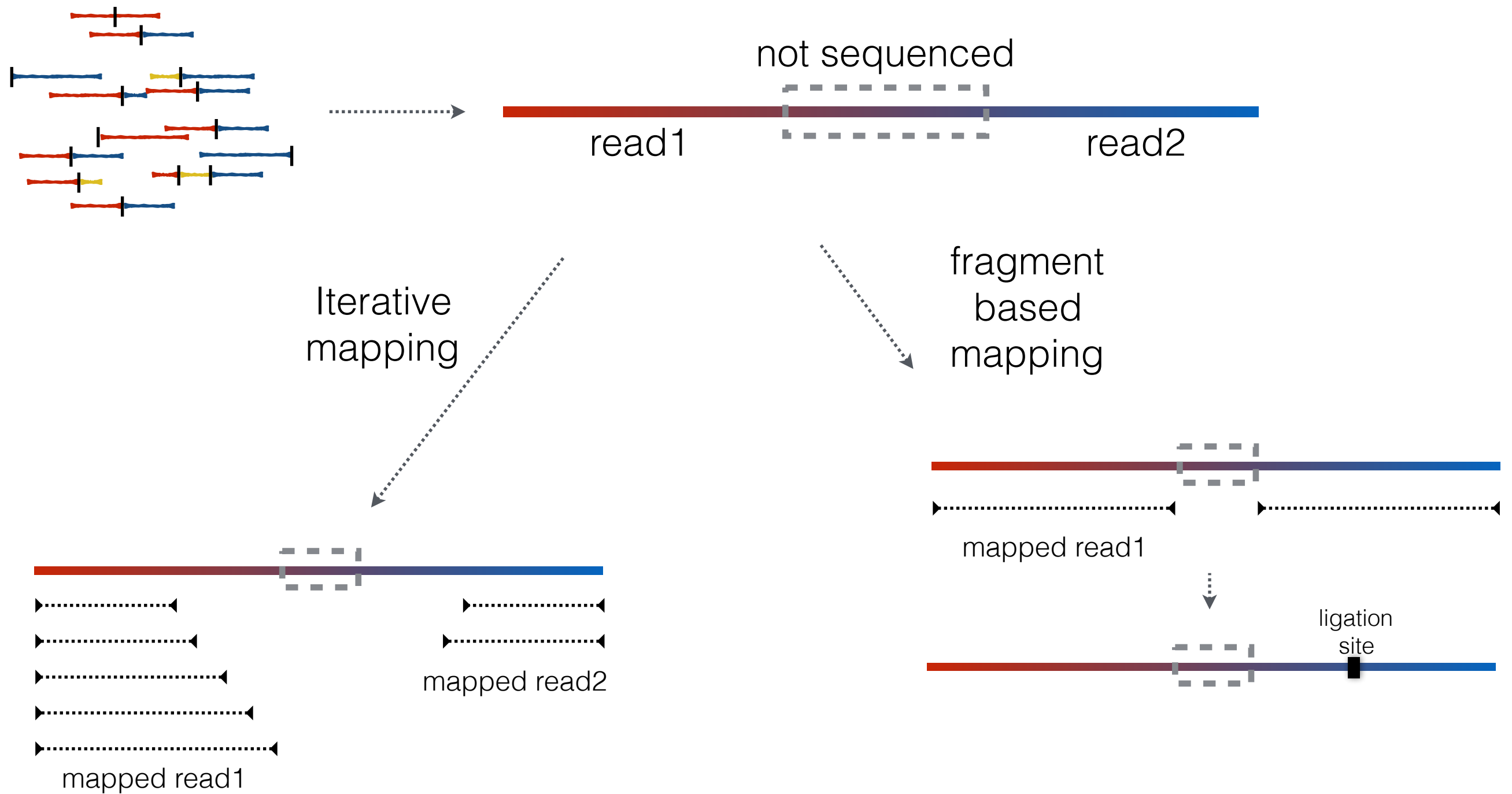
Mapping @TADbit

Serra et al. (2017). PLOS CompBio



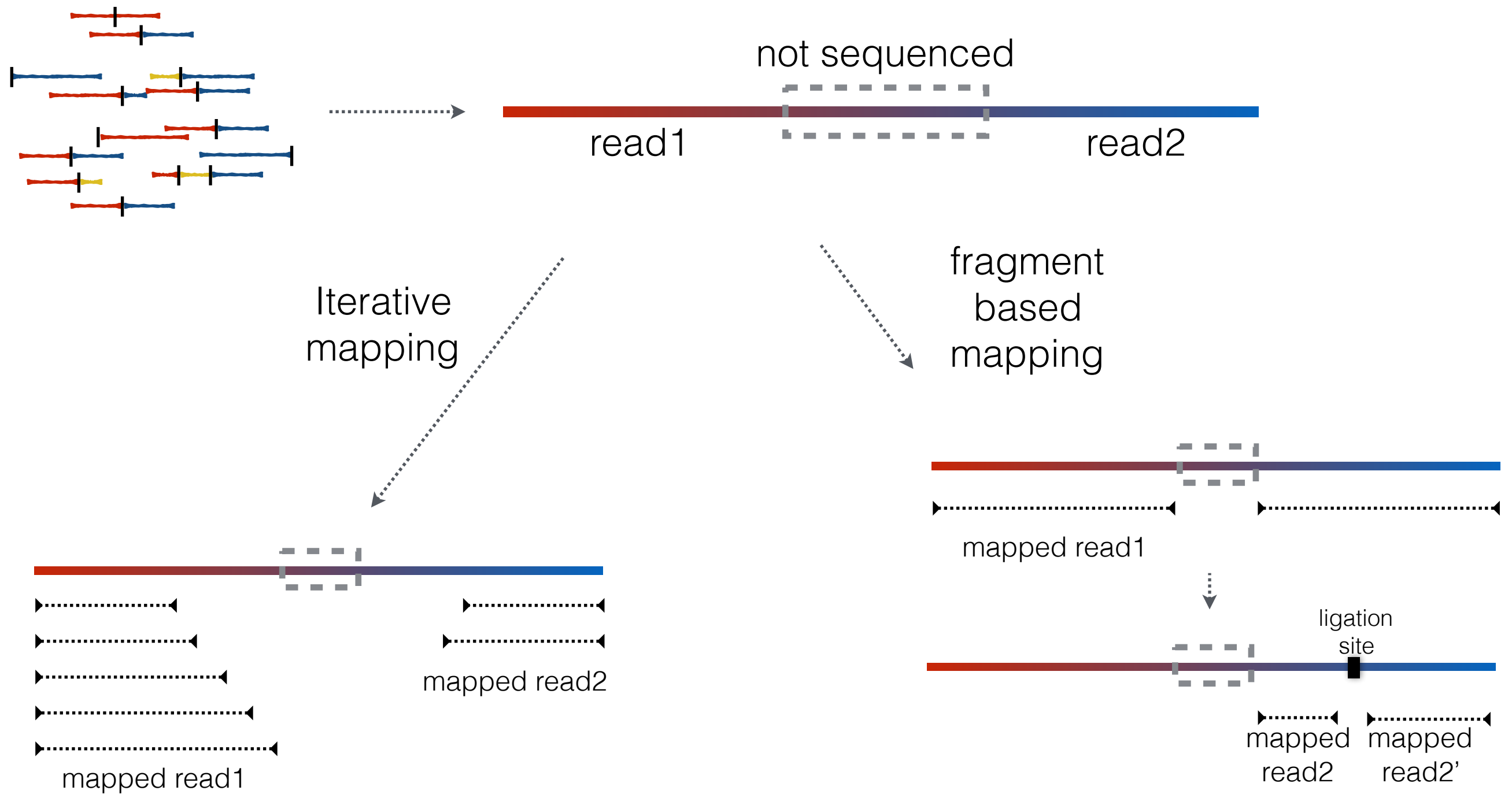
Mapping @TADbit

Serra et al. (2017). PLOS CompBio



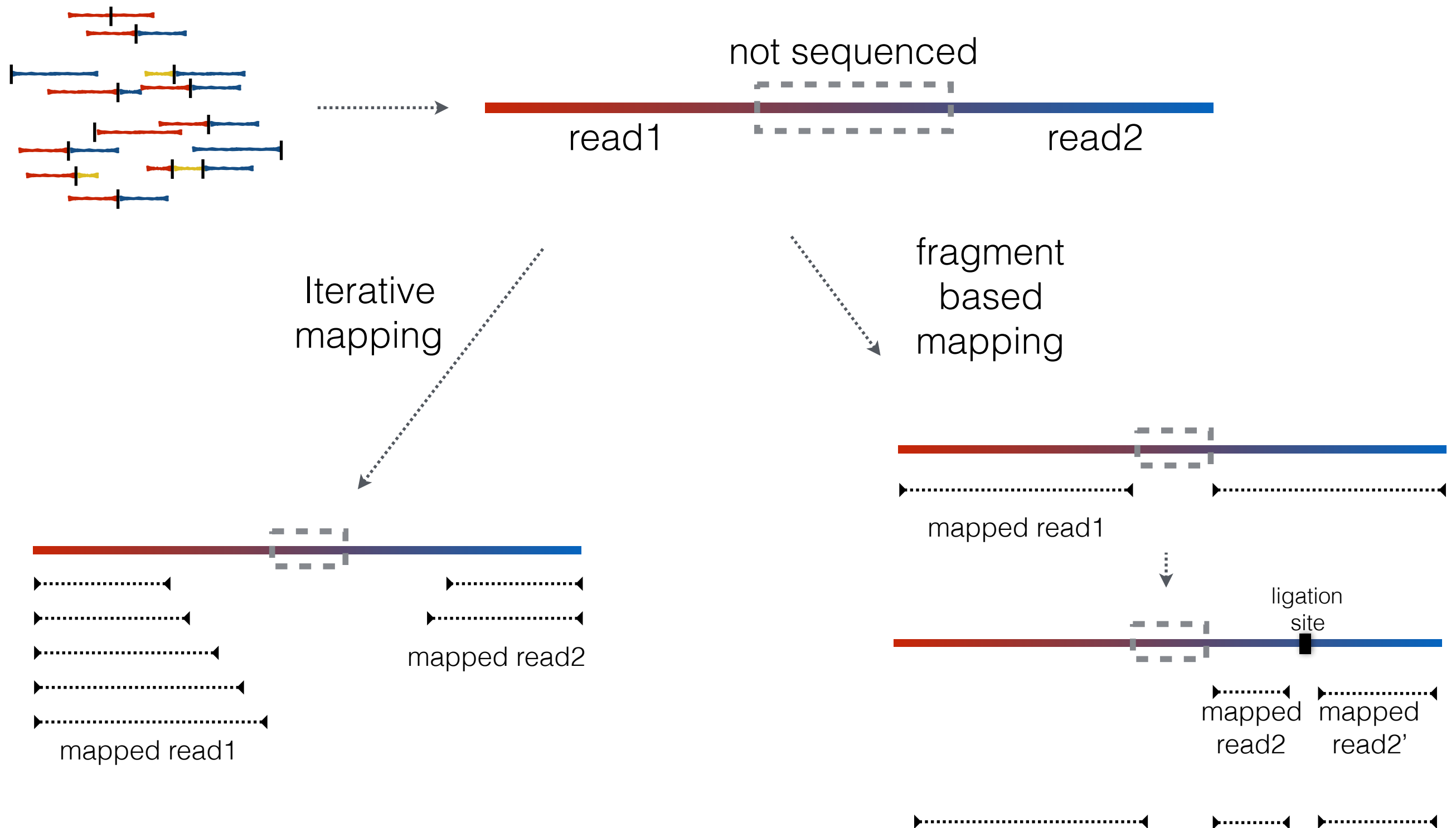
Mapping @TADbit

Serra et al. (2017). PLOS CompBio



Mapping @TADbit

Serra et al. (2017). PLOS CompBio



How much you normally map?

- 80-90% each end => 60-80% intersection
- ~1% multiple contacts
- Many of intersecting pairs will be lost in filtering...
- Final 40-60% of valid pairs
- One measure of quality is the CIS/TRANS ration (70-80% good)

<https://tinyurl.com/dir-upf-tutorial-3DG>

PRE TUTORIAL

1. DOWNLOAD DATA

Go to <https://tinyurl.com/tutorial-3DG>

2. UNTAR DATA

```
>tar -xvf tutorial_3DG.tar.gz
```

3. SET GEM PATH

```
>PATH=$PATH:/home/uXXXXX/tutorial_3DG/gem
```

TUTORIAL

1. ENTER THE TUTORIAL DIR

```
>cd tutorial_3DG
```

2. QUALITY PLOTS (seconds each command)

```
>tadbit map results --fastq FASTQs/Control_rep4.1_10M.fastq --read 1 --index db/dm6.gem --renz DpnII --qc_plot  
>tadbit map results --fastq FASTQs/Control_rep4.2_10M.fastq --read 2 --index db/dm6.gem --renz DpnII --qc_plot
```

2. MAP (~8 minutes each command, ~15 minutes all)

```
>tadbit map results --fastq FASTQs/Control_rep4.1_10M.fastq --read 1 --index db/dm6.gem --renz DpnII  
>tadbit map results --fastq FASTQs/Control_rep4.2_10M.fastq --read 2 --index db/dm6.gem --renz DpnII
```

3. PARSE (~5 minutes)

```
>tadbit parse results --compress_input --genome db/dm6.fa
```

4. FILTER (~4 minutes)

```
>tadbit filter results
```

4. NORMALIZE @100Kb and @10Kb (~2 minutes)

```
>tadbit normalize results -r 100000 --min_count 100
```

5. SEGMENT THE GENOME @100Kb A/B + TADs (~2 minutes)

```
>tadbit segment results -r 100000 --fasta db/dm6.fa -c 2L 2R 3L 3R 4 X
```

7. MODEL (~20 minutes)

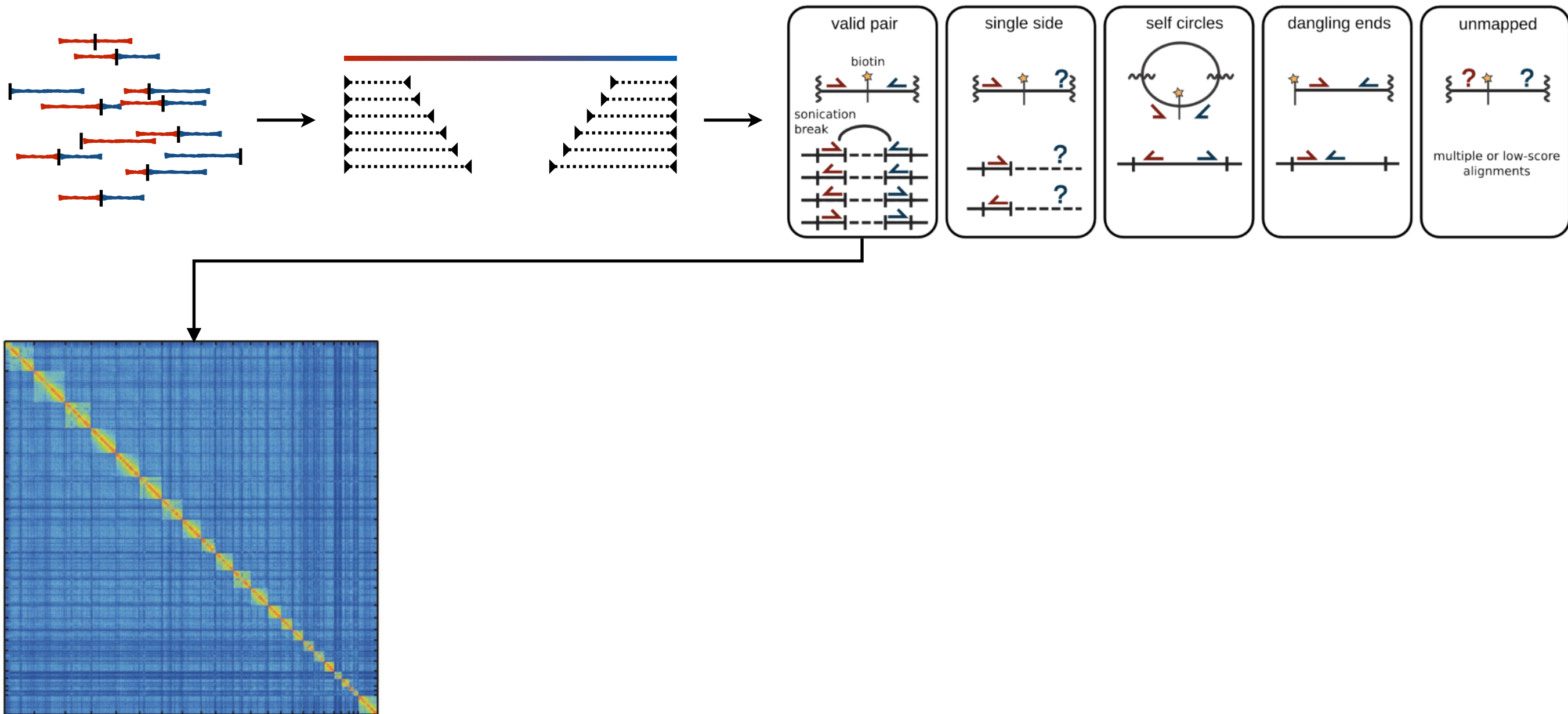
[PROBLEM: cPicke module not installed... we will walk through the process and see the final result]

```
>Model_and_analyze.py -cnf modeling.cnf
```

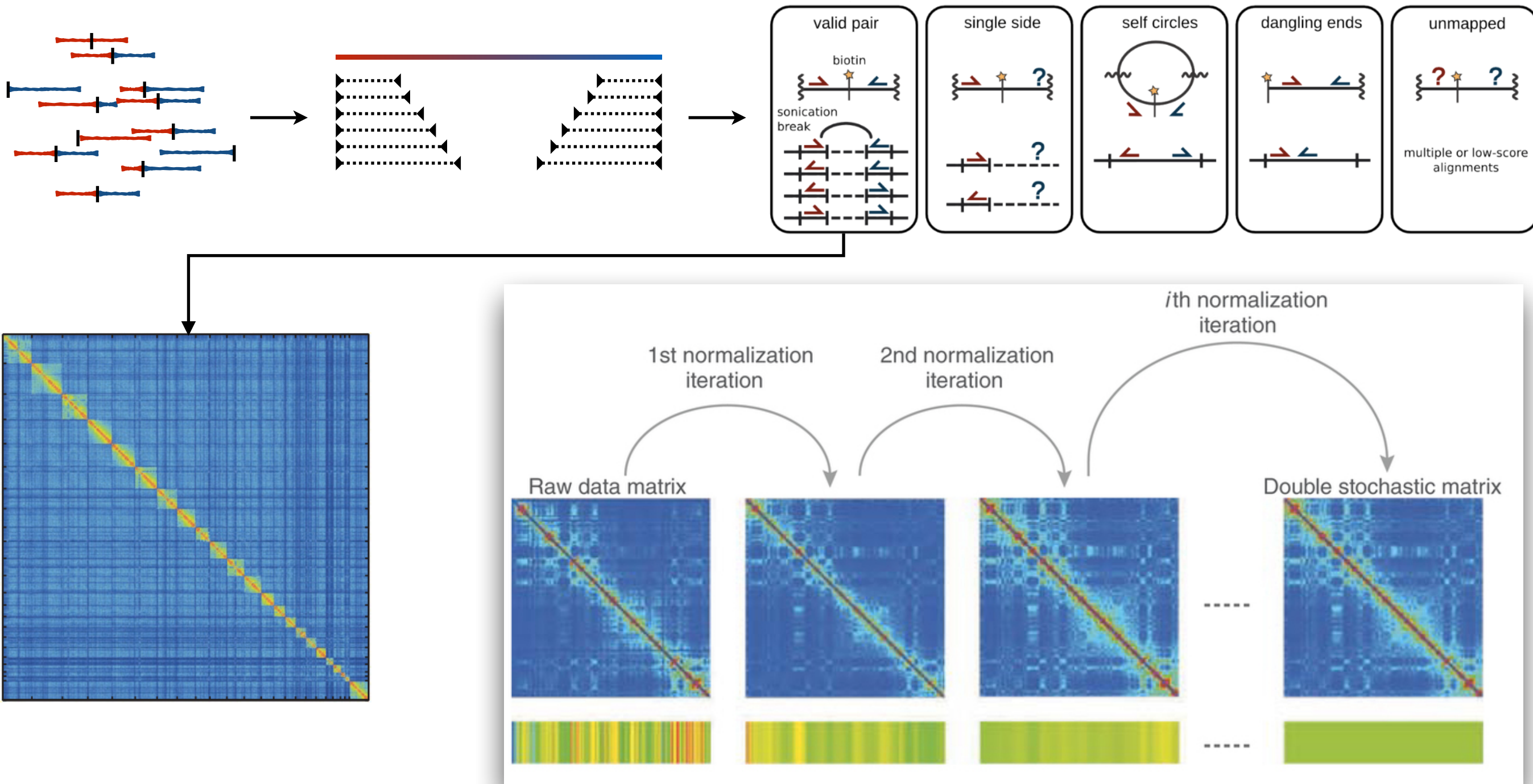


Got mapped
reads?

Interaction matrices

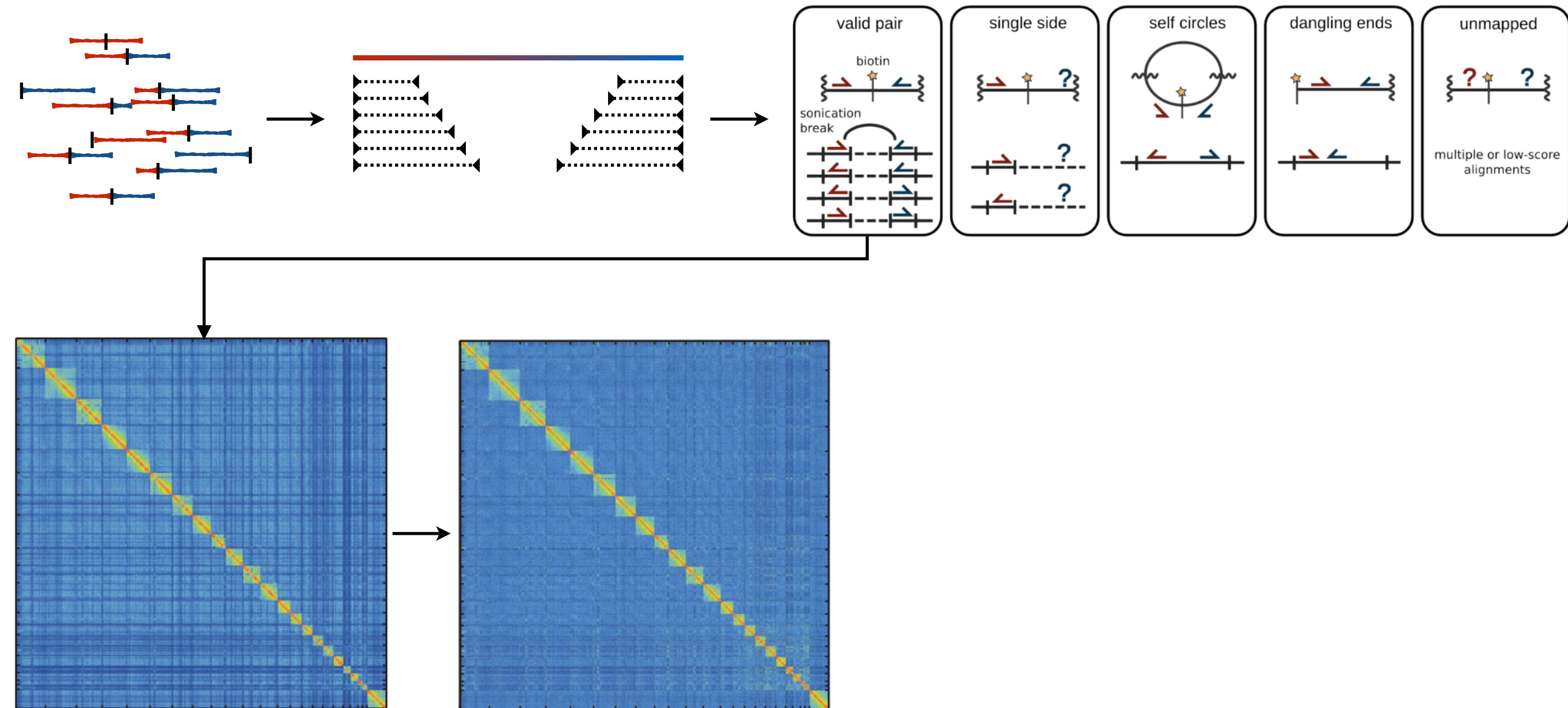


Interaction matrices

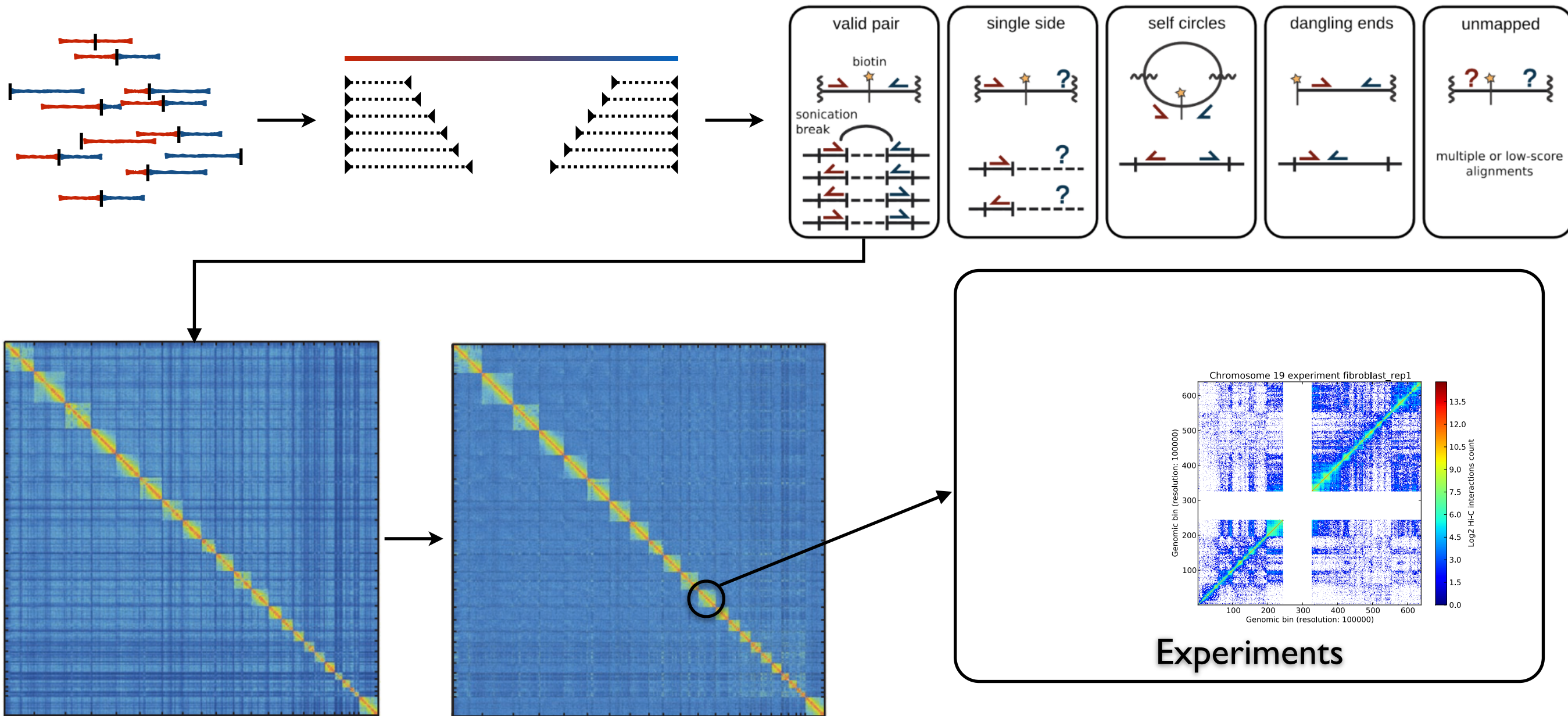


Zooming in on genome organization.
Zhou, X. J., & Alber, F. Nature Methods (2012)

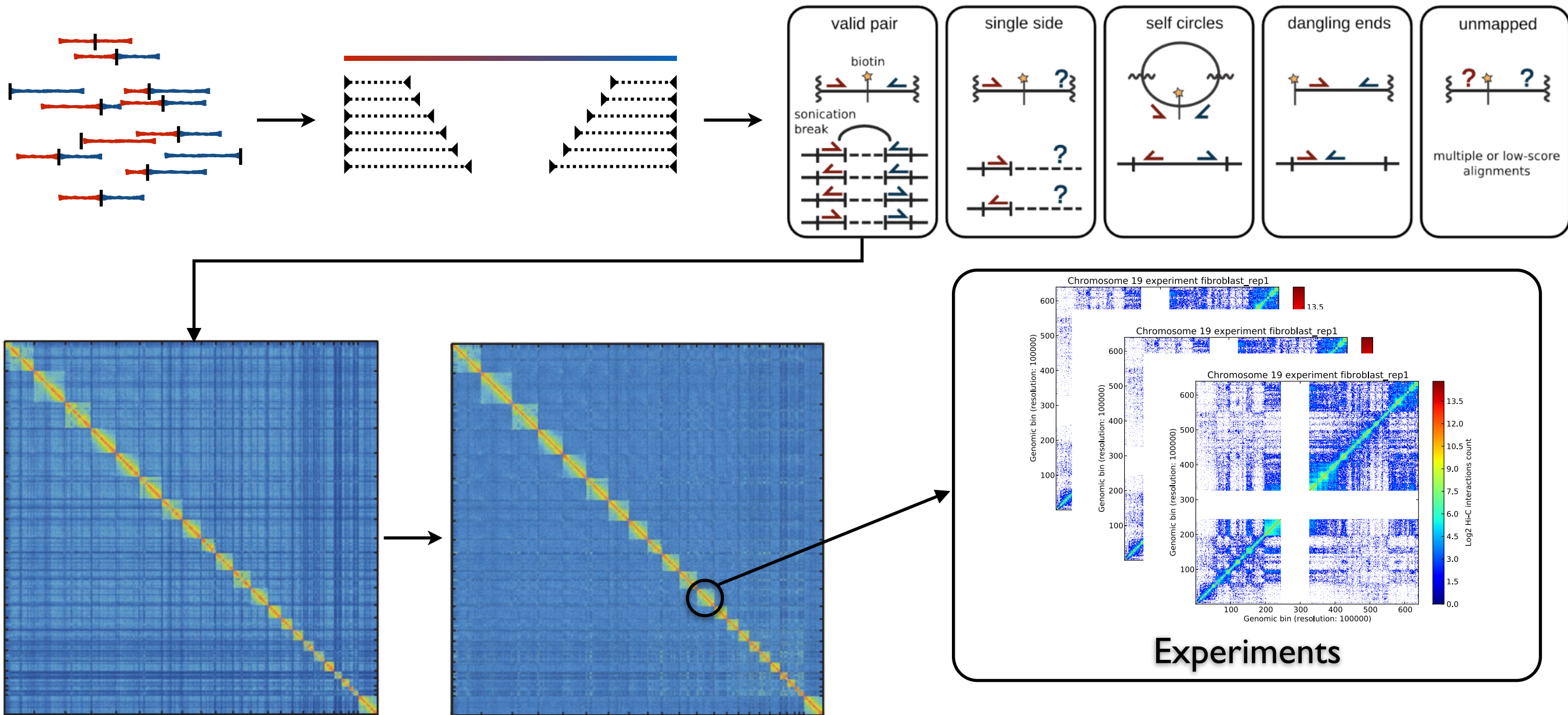
Interaction matrices



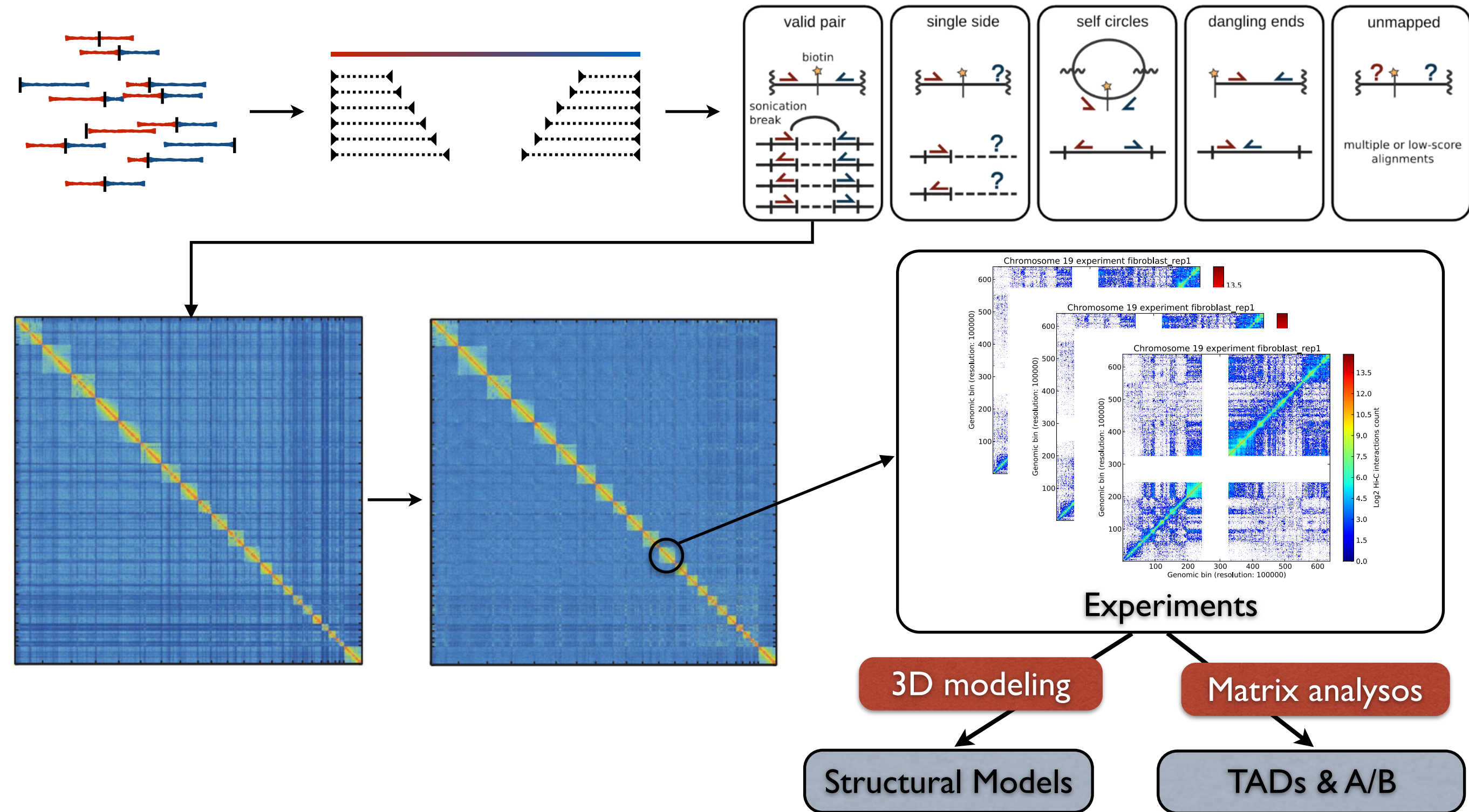
Interaction matrices



Interaction matrices



Interaction matrices



<https://tinyurl.com/dir-upf-tutorial-3DG>

PRE TUTORIAL

1. DOWNLOAD DATA

Go to <https://tinyurl.com/tutorial-3DG>

2. UNTAR DATA

```
>tar -xvf tutorial_3DG.tar.gz
```

3. SET GEM PATH

```
>PATH=$PATH:/home/uXXXXX/tutorial_3DG/gem
```

TUTORIAL

1. ENTER THE TUTORIAL DIR

```
>cd tutorial_3DG
```

2. QUALITY PLOTS (seconds each command)

```
>tadbit map results --fastq FASTQs/Control_rep4.1_10M.fastq --read 1 --index db/dm6.gem --renz DpnII --qc_plot  
>tadbit map results --fastq FASTQs/Control_rep4.2_10M.fastq --read 2 --index db/dm6.gem --renz DpnII --qc_plot
```

2. MAP (~8 minutes each command, ~15 minutes all)

```
>tadbit map results --fastq FASTQs/Control_rep4.1_10M.fastq --read 1 --index db/dm6.gem --renz DpnII  
>tadbit map results --fastq FASTQs/Control_rep4.2_10M.fastq --read 2 --index db/dm6.gem --renz DpnII
```

3. PARSE (~5 minutes)

```
>tadbit parse results --compress_input --genome db/dm6.fa
```

4. FILTER (~4 minutes)

```
>tadbit filter results
```

4. NORMALIZE @100Kb and @10Kb (~2 minutes)

```
>tadbit normalize results -r 100000 --min_count 100
```

5. SEGMENT THE GENOME @100Kb A/B + TADs (~2 minutes)

```
>tadbit segment results -r 100000 --fasta db/dm6.fa -c 2L 2R 3L 3R 4 X
```

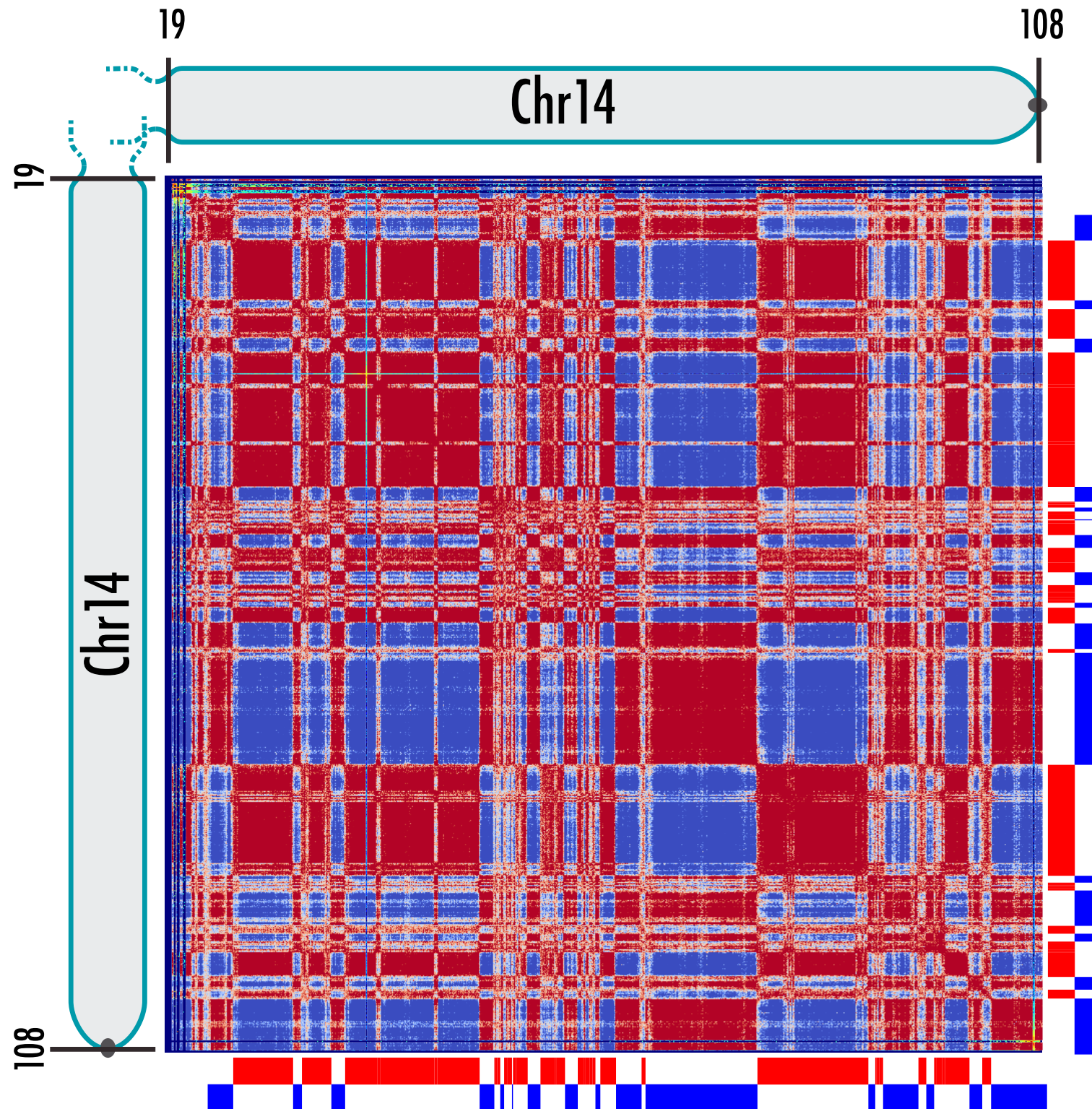
7. MODEL (~20 minutes)

[PROBLEM: cPicke module not installed... we will walk through the process and see the final result]

```
>Model_and_analyze.py -cnf modeling.cnf
```

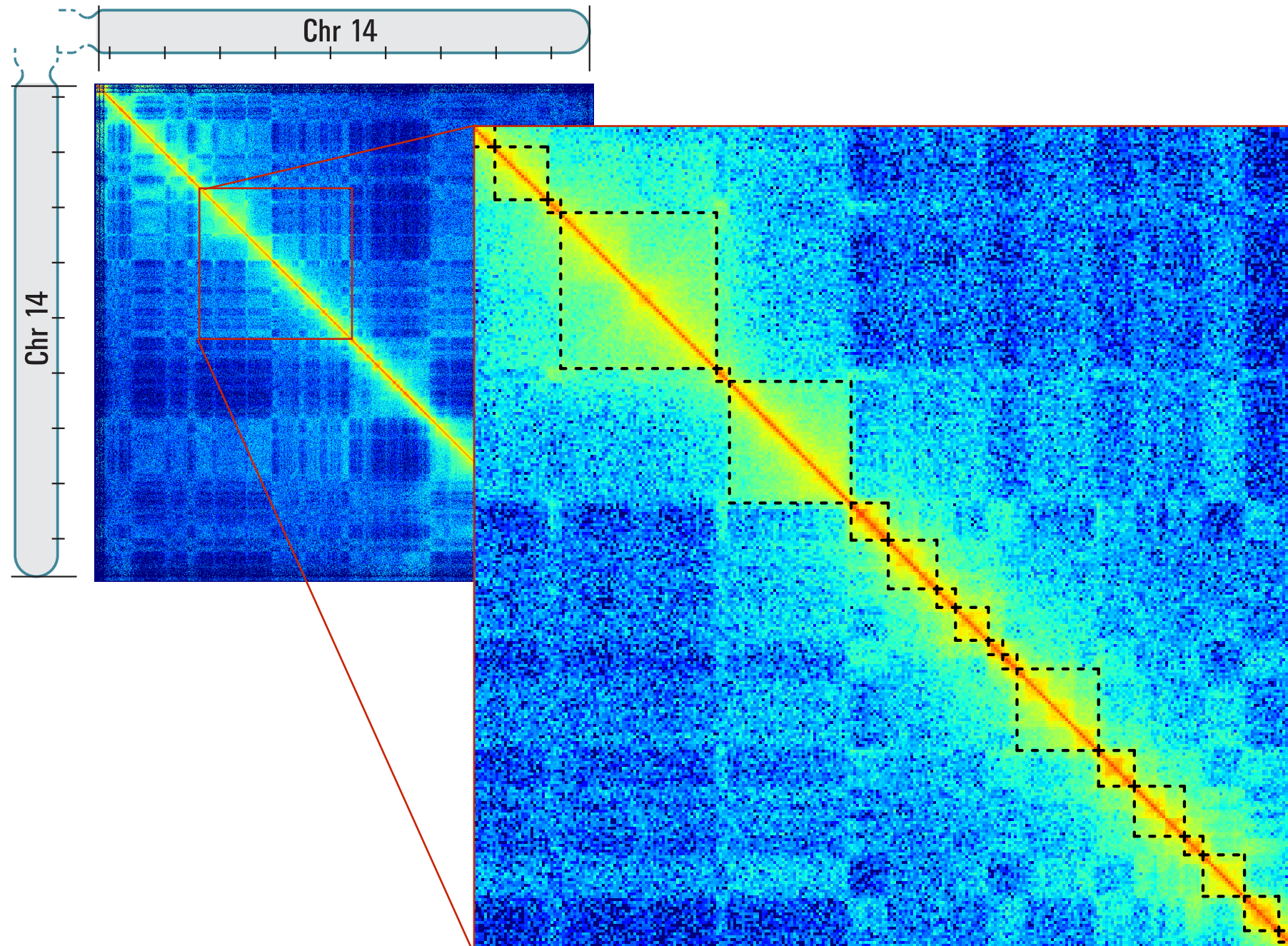
A/B Compartment

Chromosome 14



TADs

Chromosome 14



<https://tinyurl.com/dir-upf-tutorial-3DG>

PRE TUTORIAL

1. DOWNLOAD DATA

Go to <https://tinyurl.com/tutorial-3DG>

2. UNTAR DATA

```
>tar -xvf tutorial_3DG.tar.gz
```

3. SET GEM PATH

```
>PATH=$PATH:/home/uXXXXX/tutorial_3DG/gem
```

TUTORIAL

1. ENTER THE TUTORIAL DIR

```
>cd tutorial_3DG
```

2. QUALITY PLOTS (seconds each command)

```
>tadbit map results --fastq FASTQs/Control_rep4.1_10M.fastq --read 1 --index db/dm6.gem --renz DpnII --qc_plot  
>tadbit map results --fastq FASTQs/Control_rep4.2_10M.fastq --read 2 --index db/dm6.gem --renz DpnII --qc_plot
```

2. MAP (~8 minutes each command, ~15 minutes all)

```
>tadbit map results --fastq FASTQs/Control_rep4.1_10M.fastq --read 1 --index db/dm6.gem --renz DpnII  
>tadbit map results --fastq FASTQs/Control_rep4.2_10M.fastq --read 2 --index db/dm6.gem --renz DpnII
```

3. PARSE (~5 minutes)

```
>tadbit parse results --compress_input --genome db/dm6.fa
```

4. FILTER (~4 minutes)

```
>tadbit filter results
```

4. NORMALIZE @100Kb and @10Kb (~2 minutes)

```
>tadbit normalize results -r 100000 --min_count 100
```

5. SEGMENT THE GENOME @100Kb A/B + TADs (~2 minutes)

```
>tadbit segment results -r 100000 --fasta db/dm6.fa -c 2L 2R 3L 3R 4 X
```

7. MODEL (~20 minutes)

[PROBLEM: cPicke module not installed... we will walk through the process and see the final result]

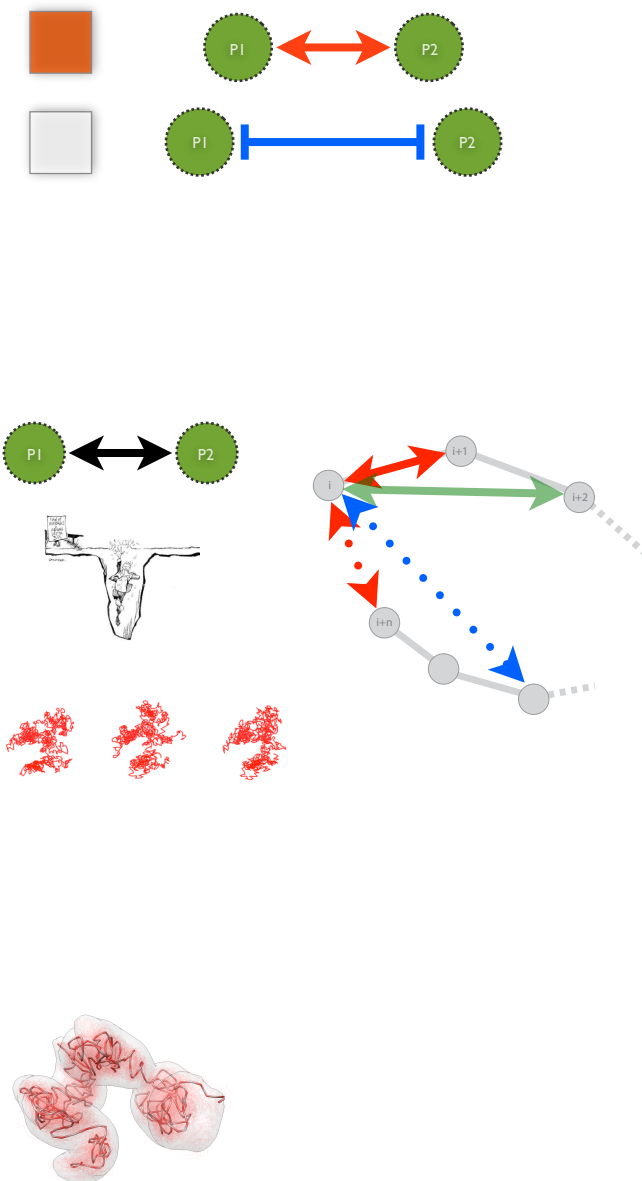
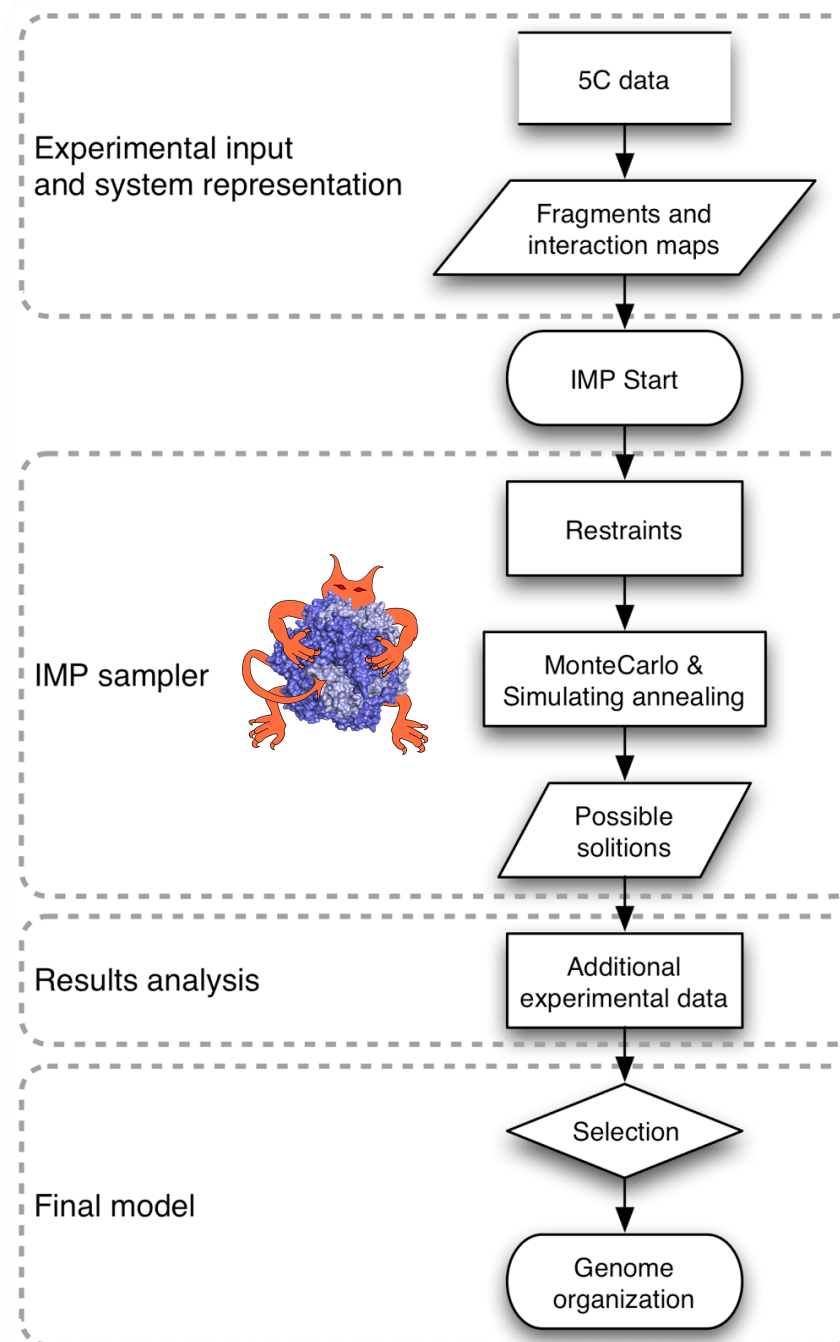
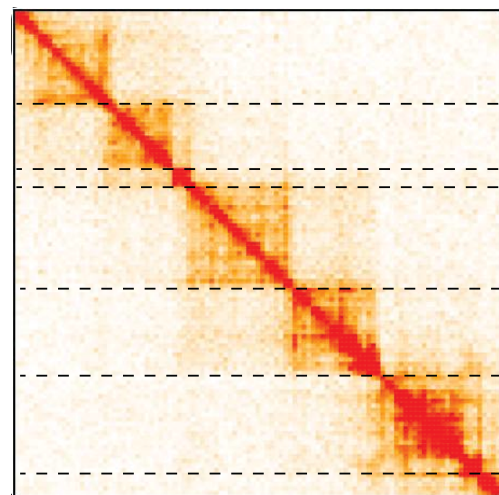
```
>Model_and_analyze.py -cnf modeling.cnf
```



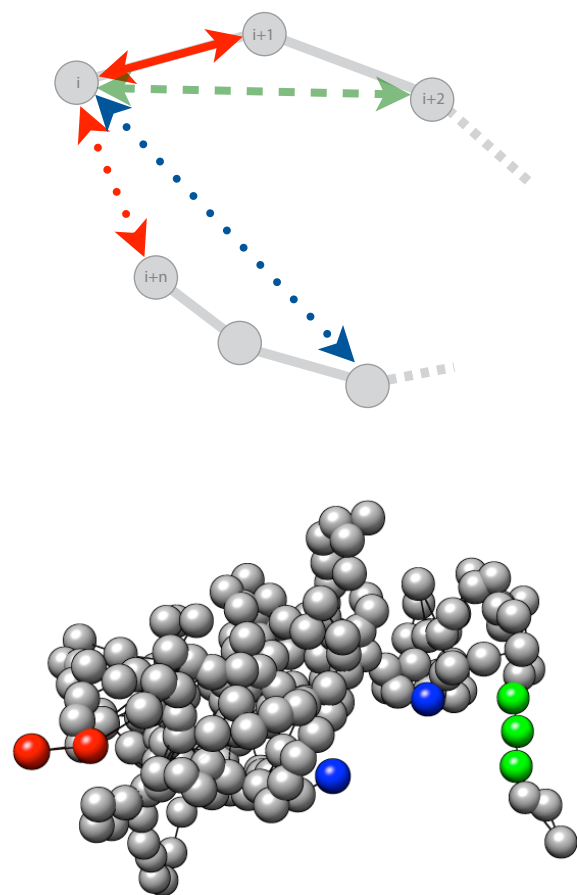

Got normalized
Hi-C maps?



<http://3DGenomes.org>
<http://www.integrativemodeling.org>



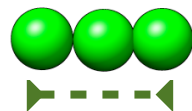
Model representation and scoring



$d = d_0$



$d < d_0$

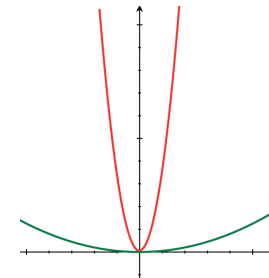


$d > d_0$



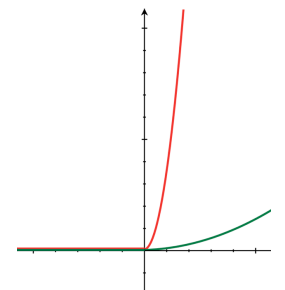
Harmonic

$$H_{i,j} = k(d_{i,j} - d_{i,j}^0)^2$$



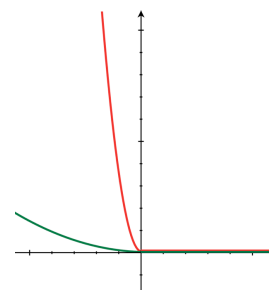
Harmonic Upper Bound

$$\begin{cases} \text{if } d_{i,j} \geq d_{i,j}^0; & ubH_{i,j} = k(d_{i,j} - d_{i,j}^0)^2 \\ \text{if } d_{i,j} < d_{i,j}^0; & ubH_{i,j} = 0 \end{cases}$$

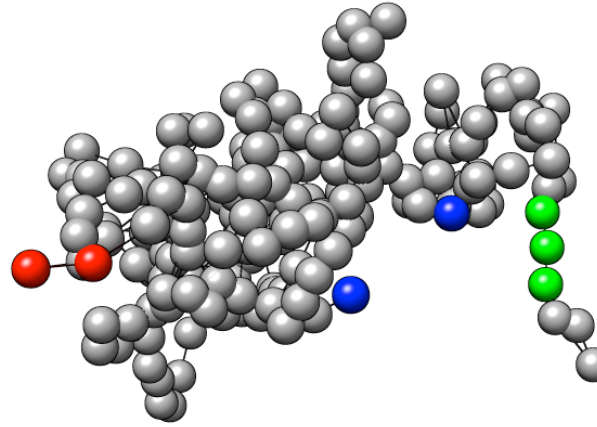


Harmonic Lower Bound

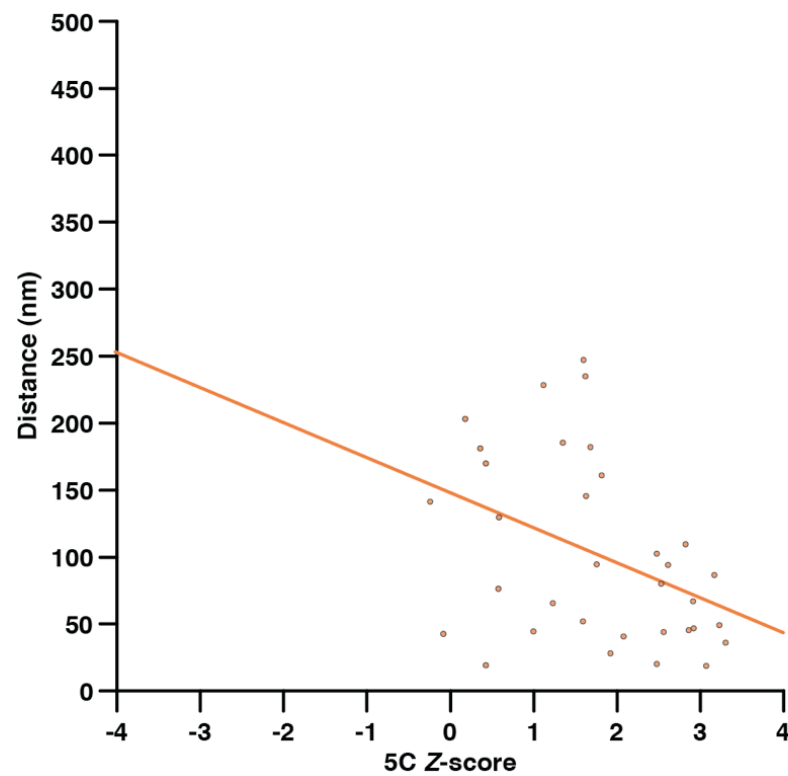
$$\begin{cases} \text{if } d_{i,j} \leq d_{i,j}^0; & lbH_{i,j} = k(d_{i,j} - d_{i,j}^0)^2 \\ \text{if } d_{i,j} > d_{i,j}^0; & lbH_{i,j} = 0 \end{cases}$$



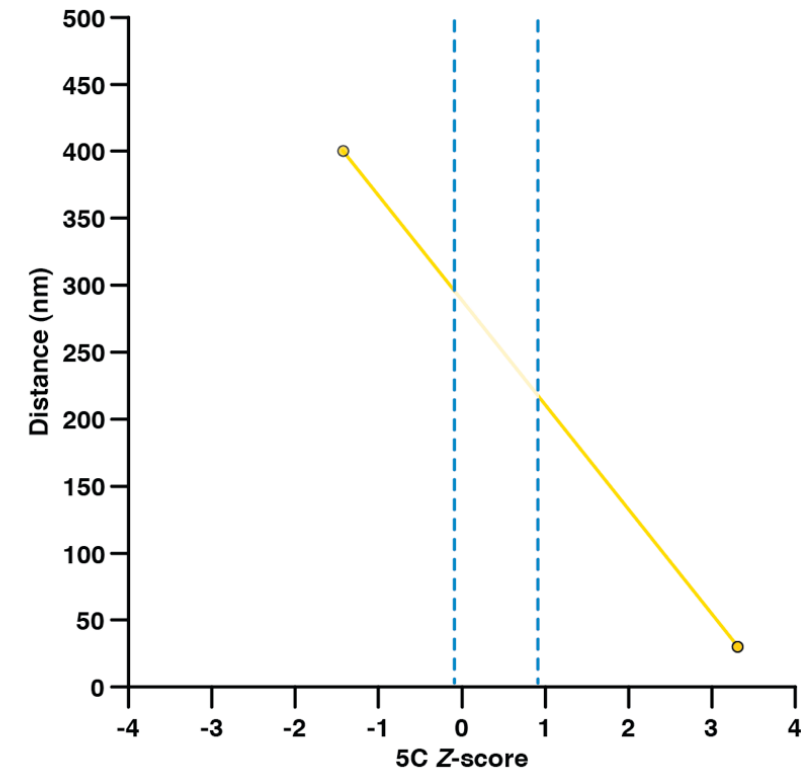
From 3C data to spatial distances



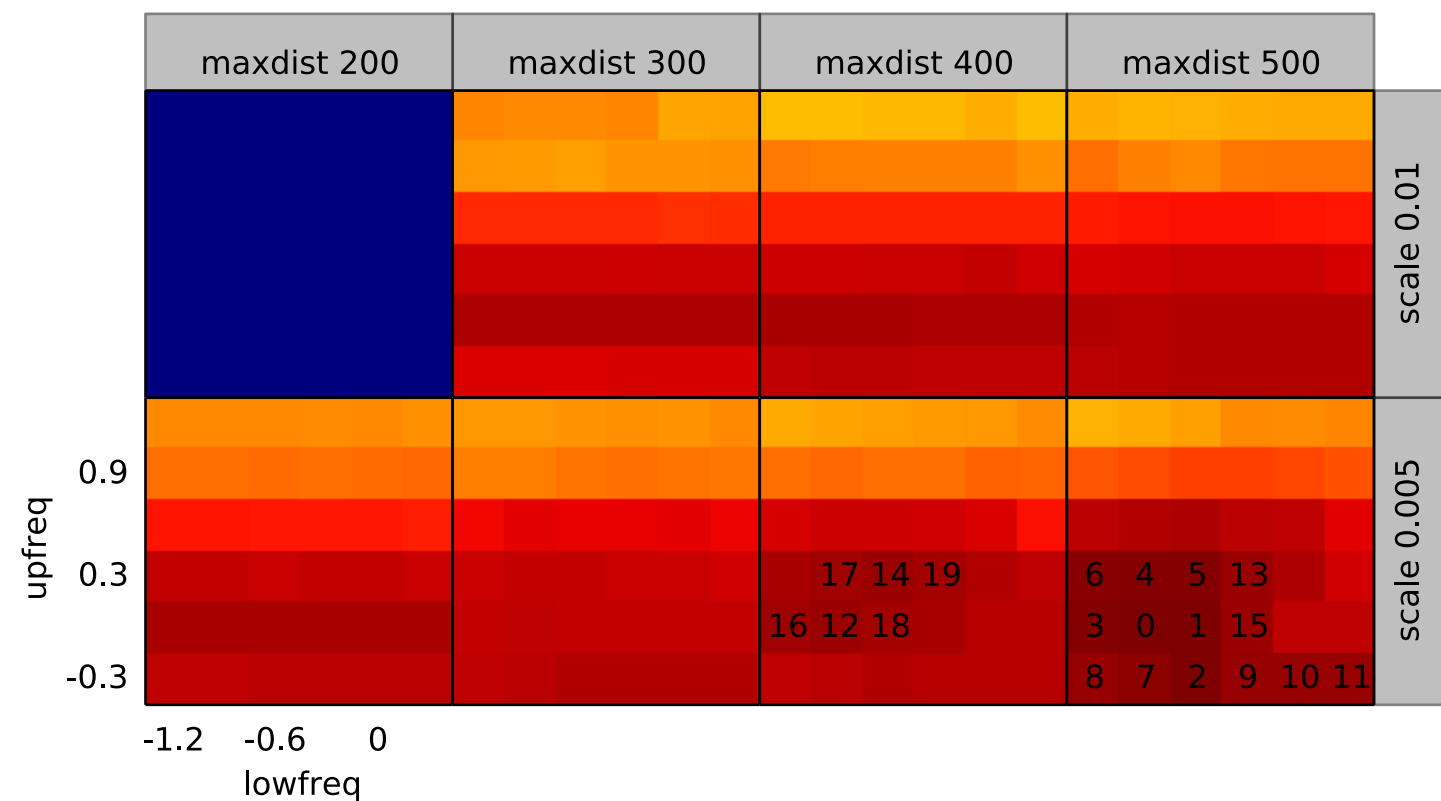
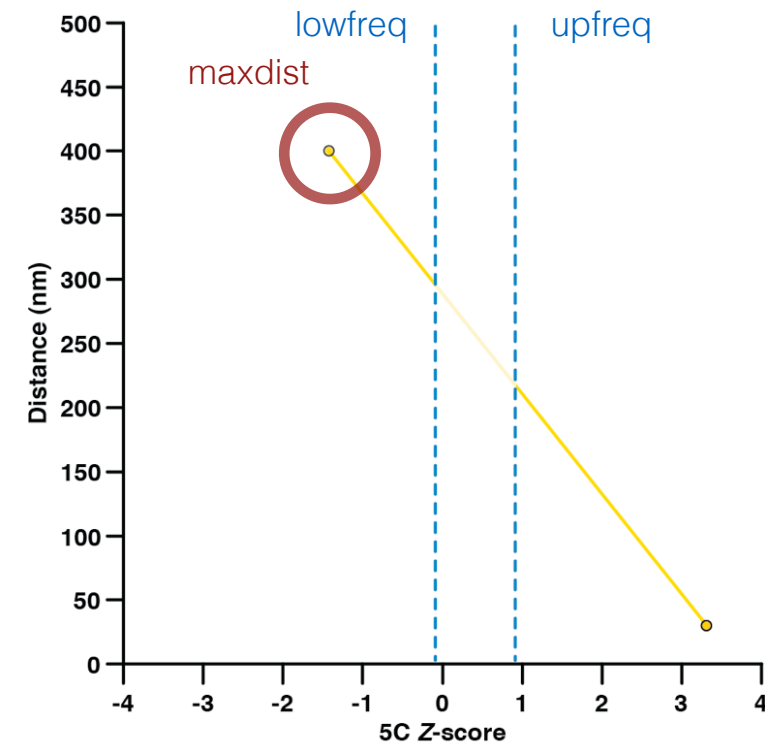
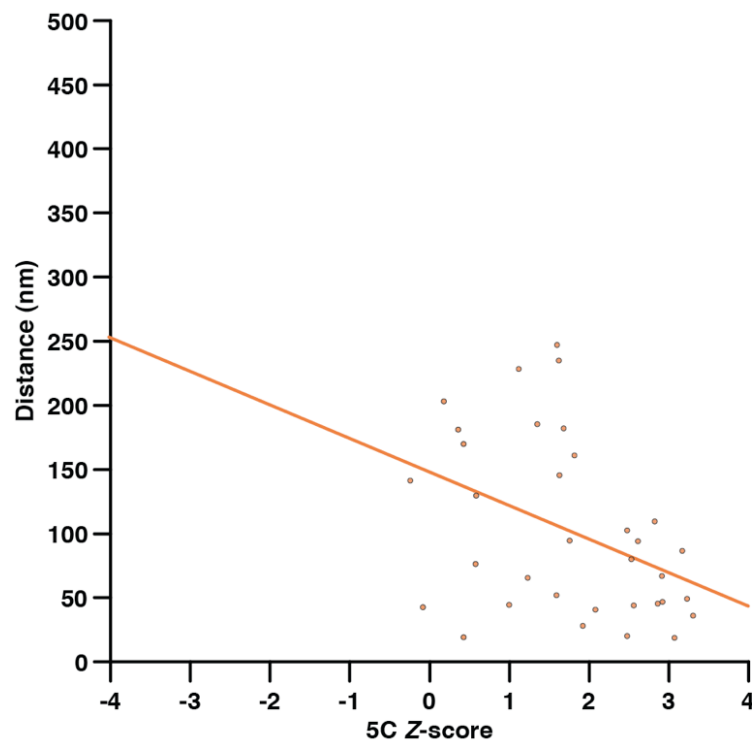
Neighbor fragments



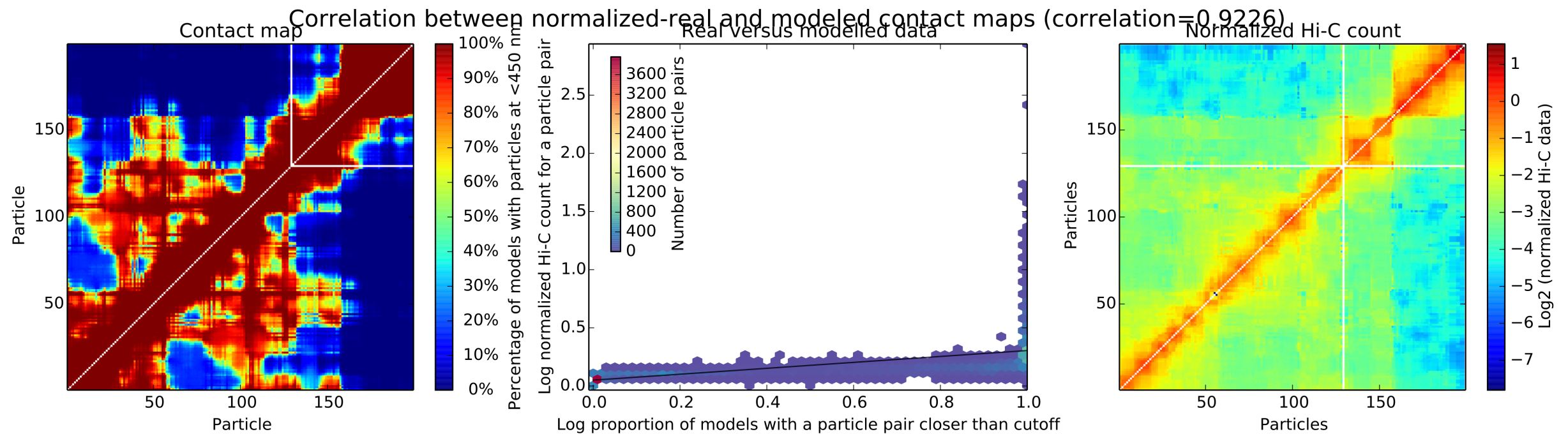
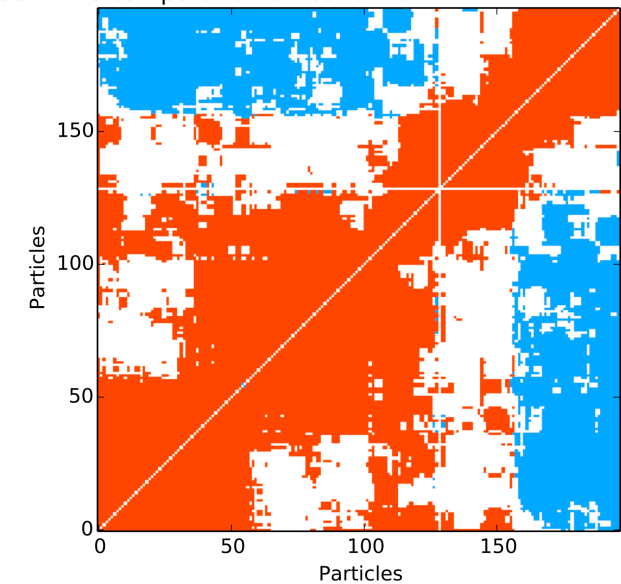
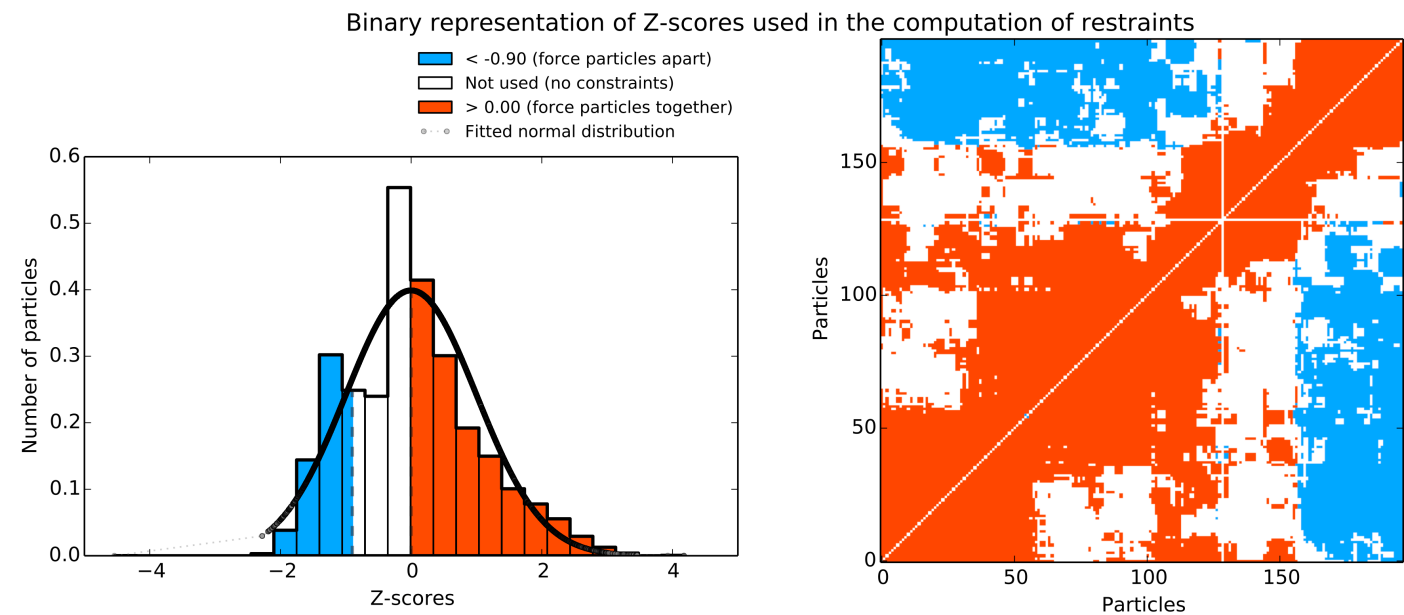
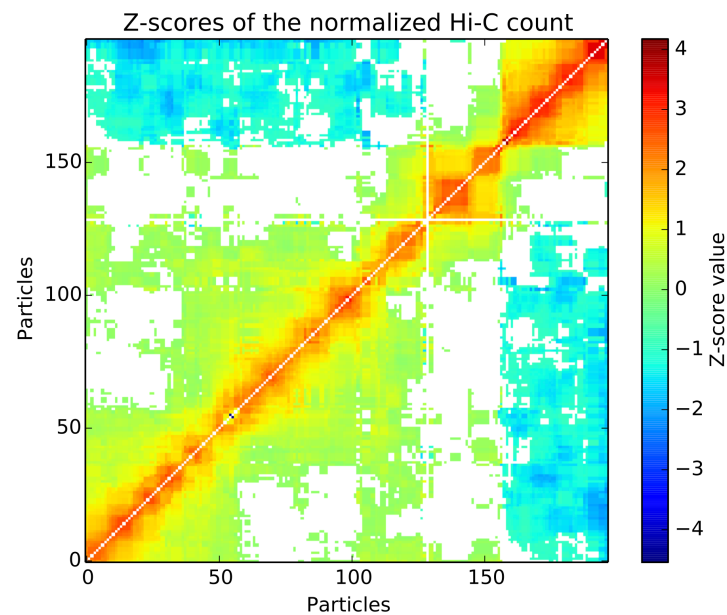
Non-Nighbor fragments



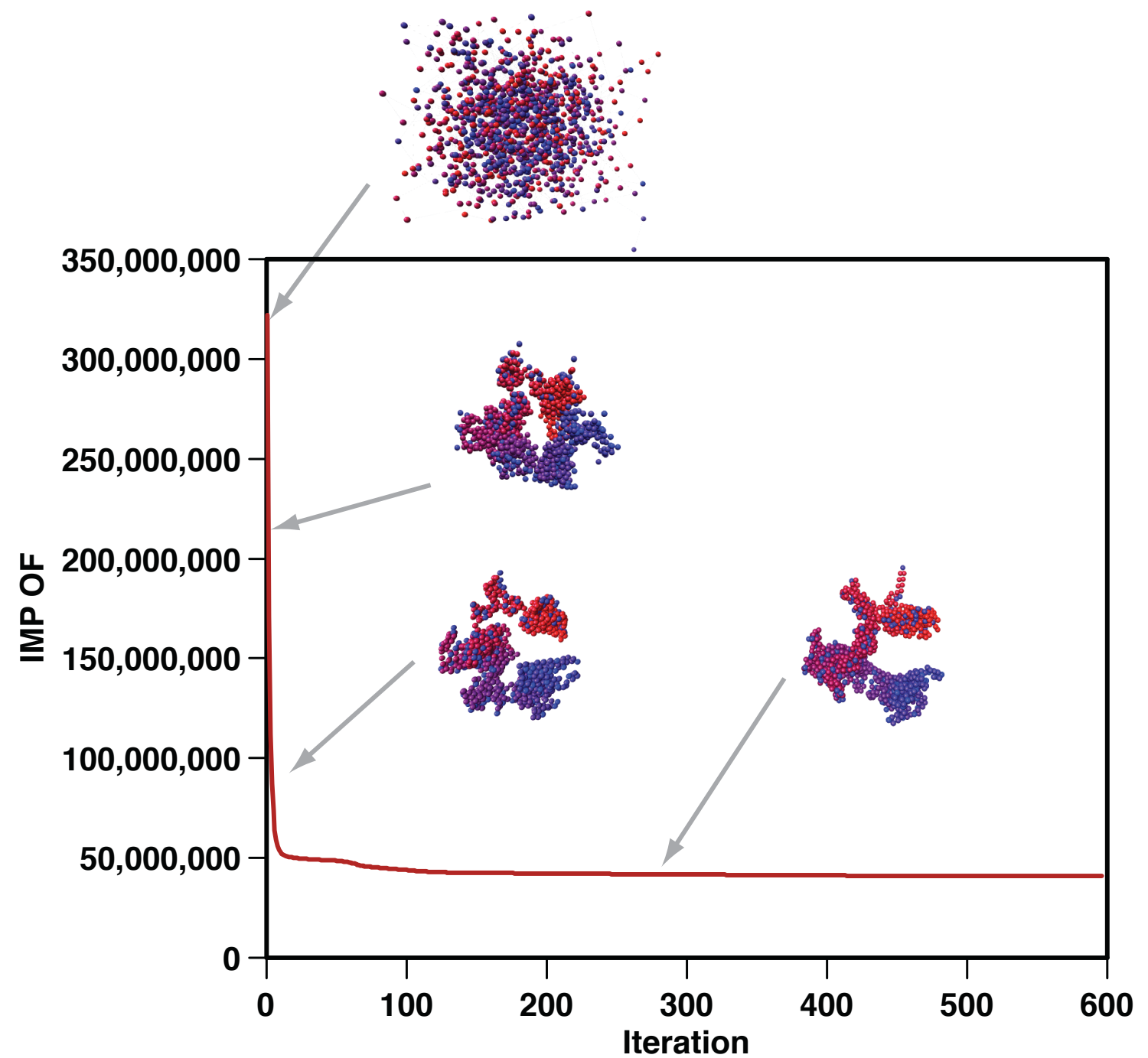
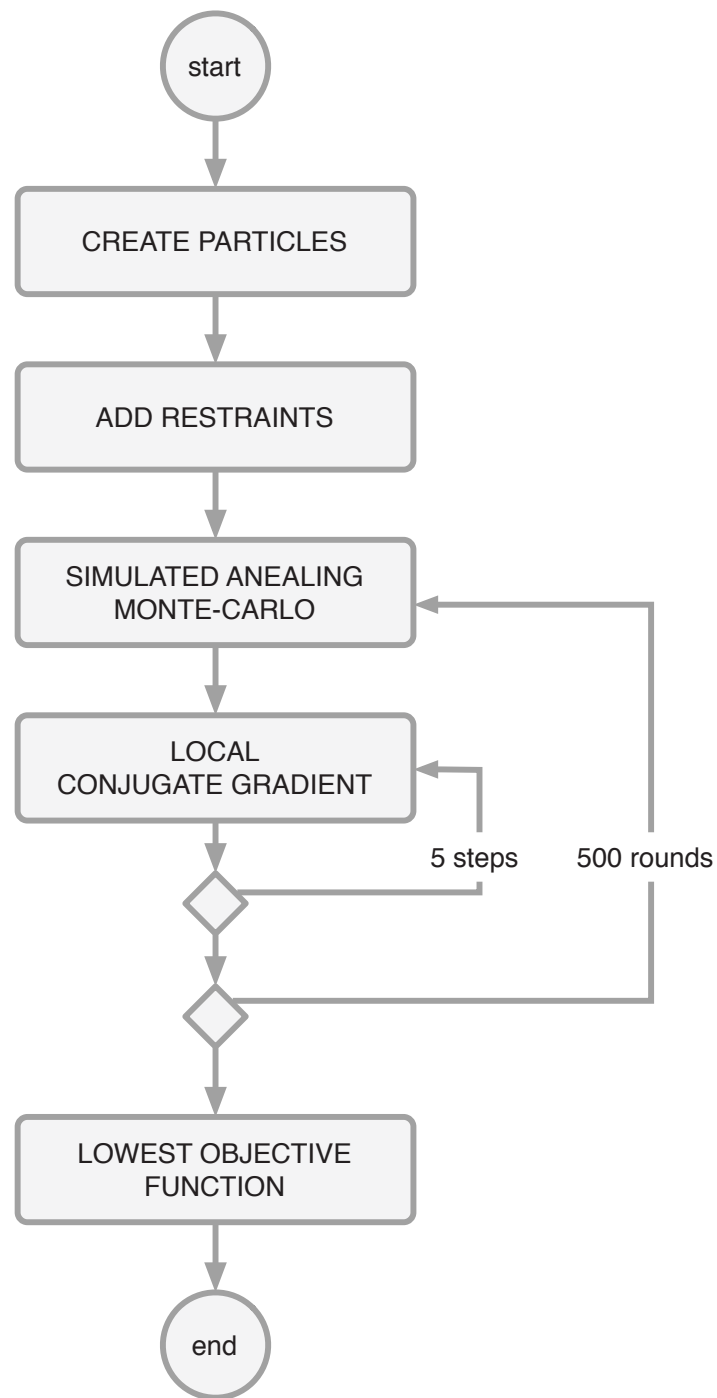
Parameter optimization



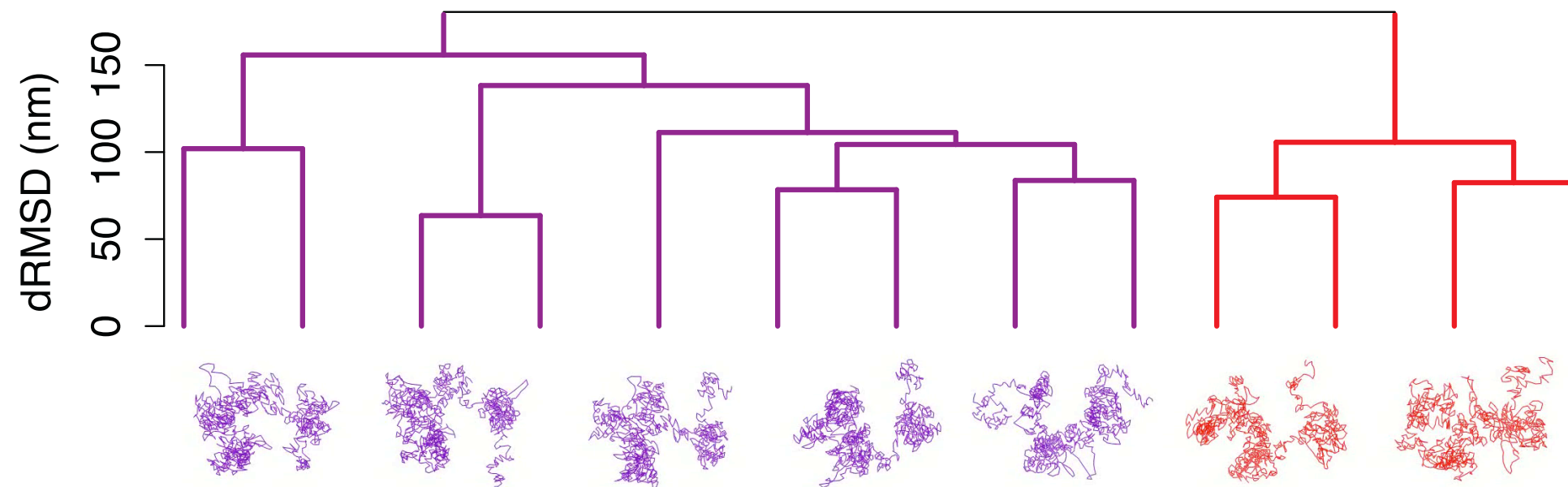
Parameter optimization



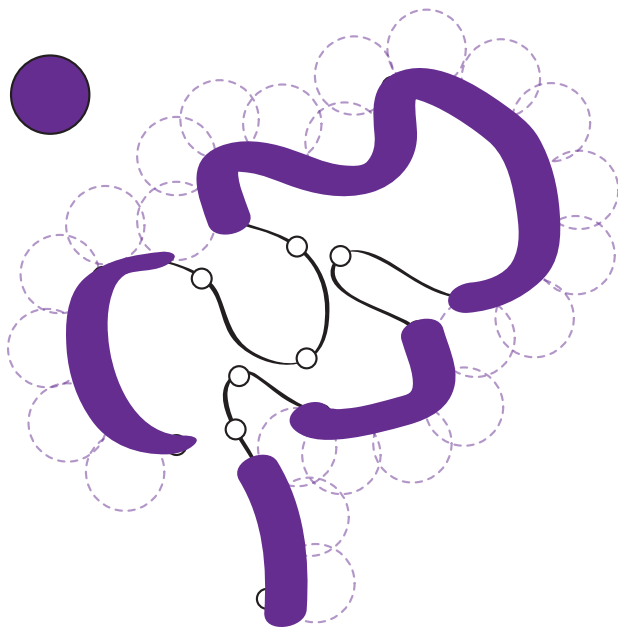
Optimization of the scoring function



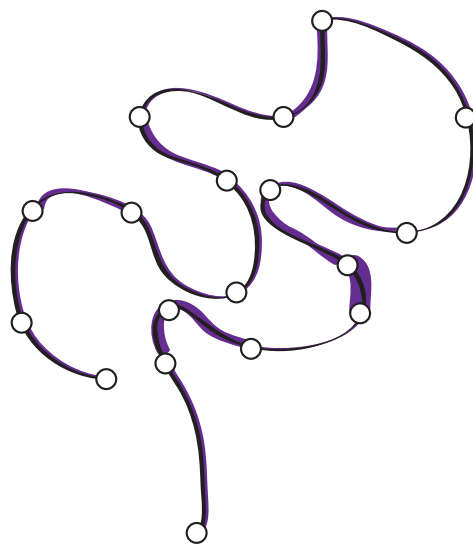
Model analysis: clustering and structural features



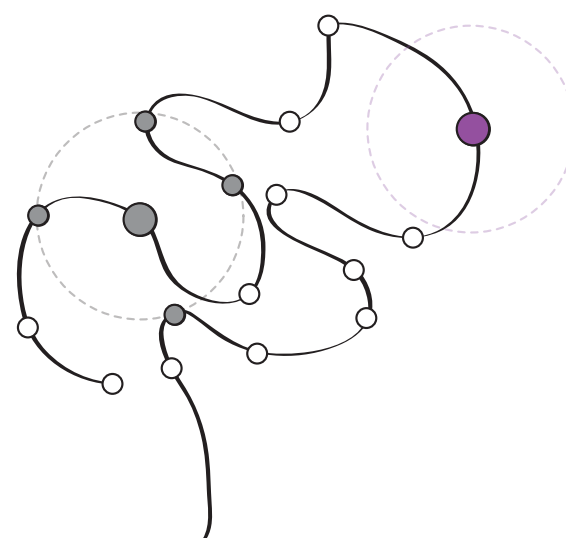
Accessibility (%)



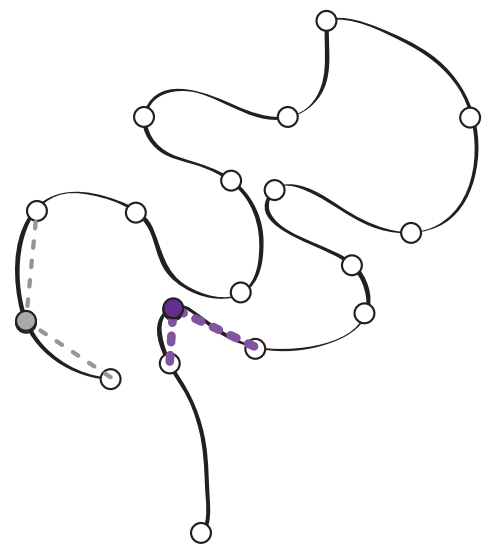
Density (bp/nm)



Interactions



Angle



<http://marciuslab.org>
<http://3DGenomes.org>
<http://cnag.crg.eu>

cnag



<http://sgt.cnag.cat/www/presentations/>

David Castillo
Yasmina Cuartero
Marco Di Stefano
Irene Farabella
Silvia Galan
Mike Goodstadt
Maria Marti-Marimon
Francesca Mugianesi
Julen Mendieta
Juan Rodriguez
Paula Soler
Aleksandra Sparavier



Multiscale
Complex
Genomics



MINISTERIO
DE ECONOMÍA, INDUSTRIA
Y COMPETITIVIDAD



**Generalitat
de Catalunya**



**Fundació
La Marató de TV3**

