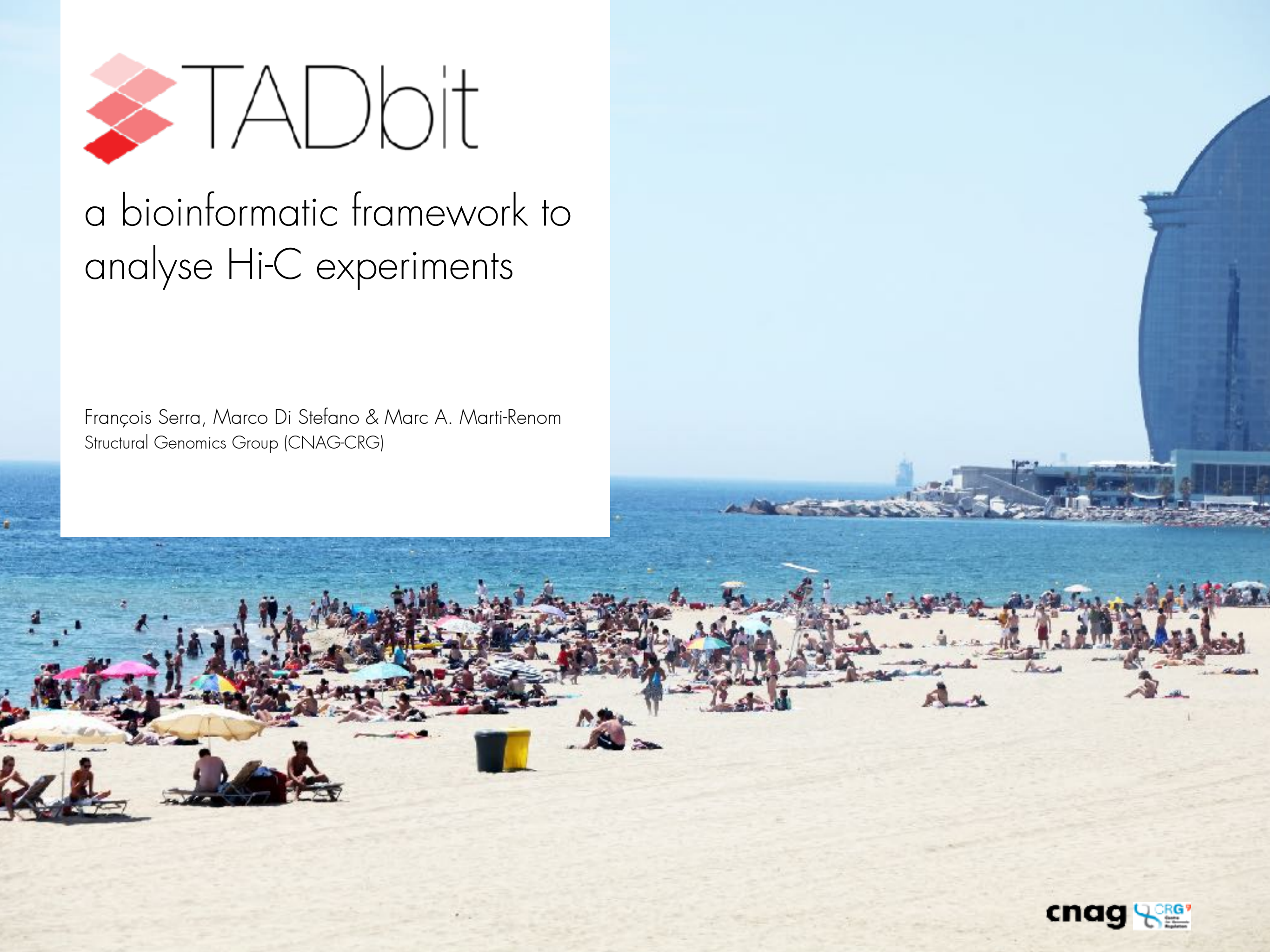


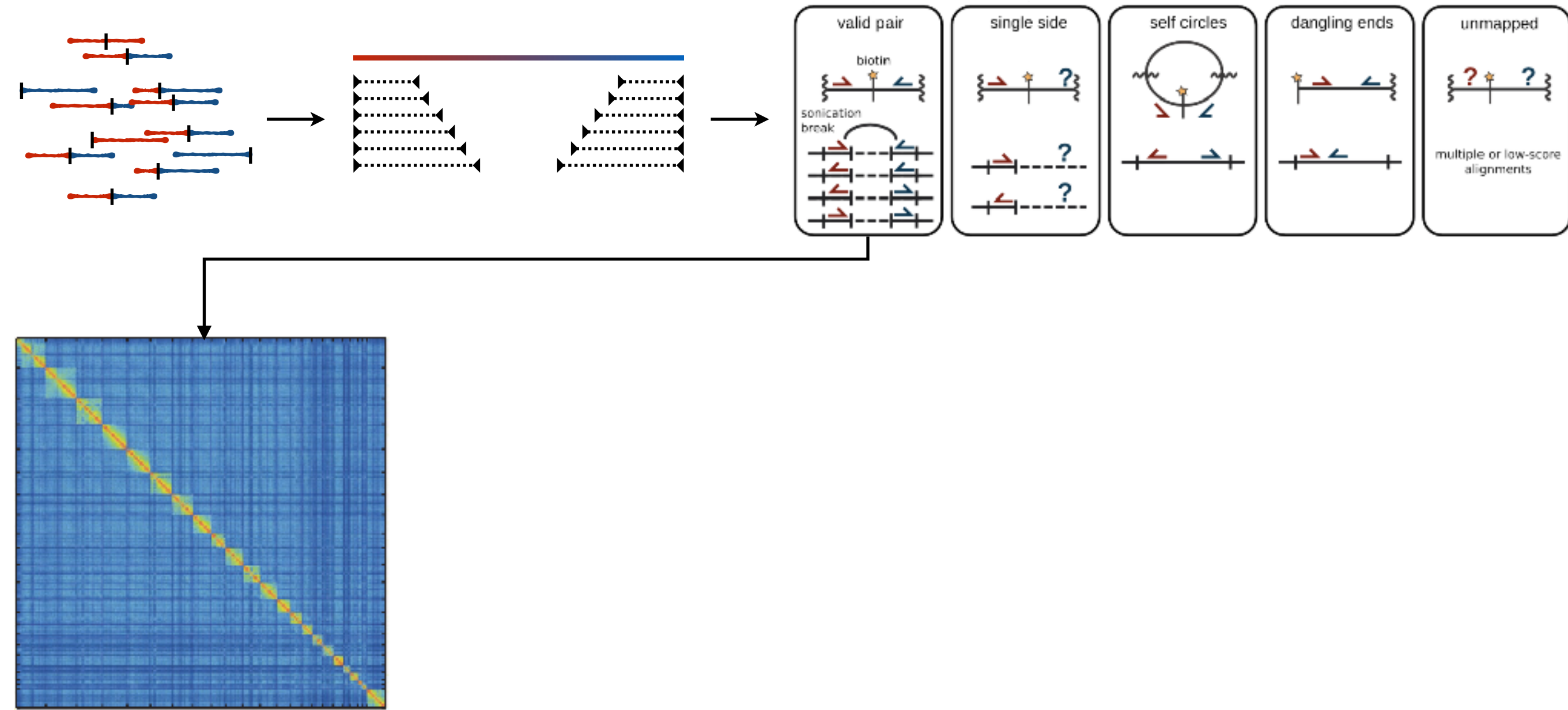


a bioinformatic framework to  
analyse Hi-C experiments

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

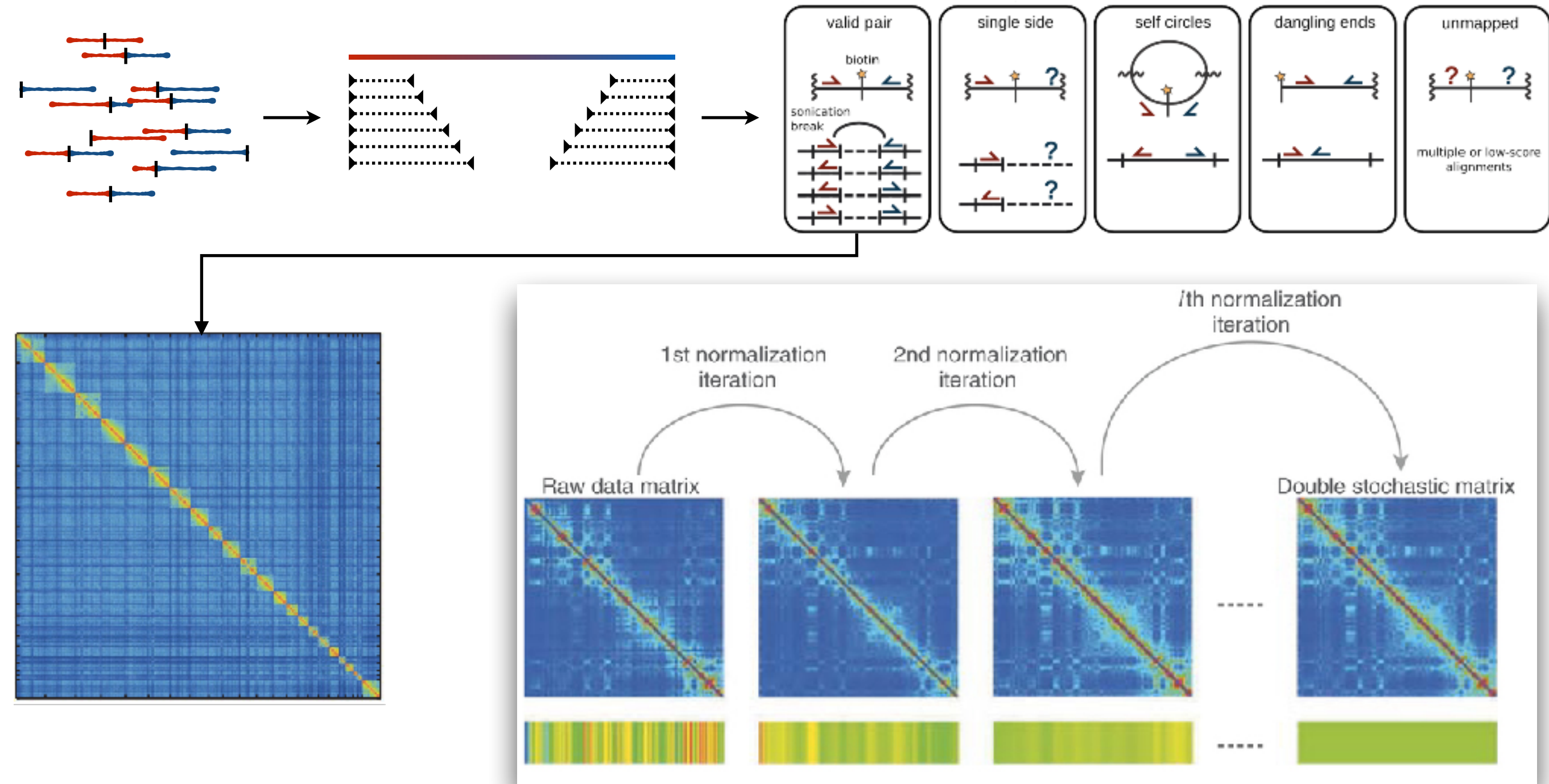


# From FASTQ to interaction matrices





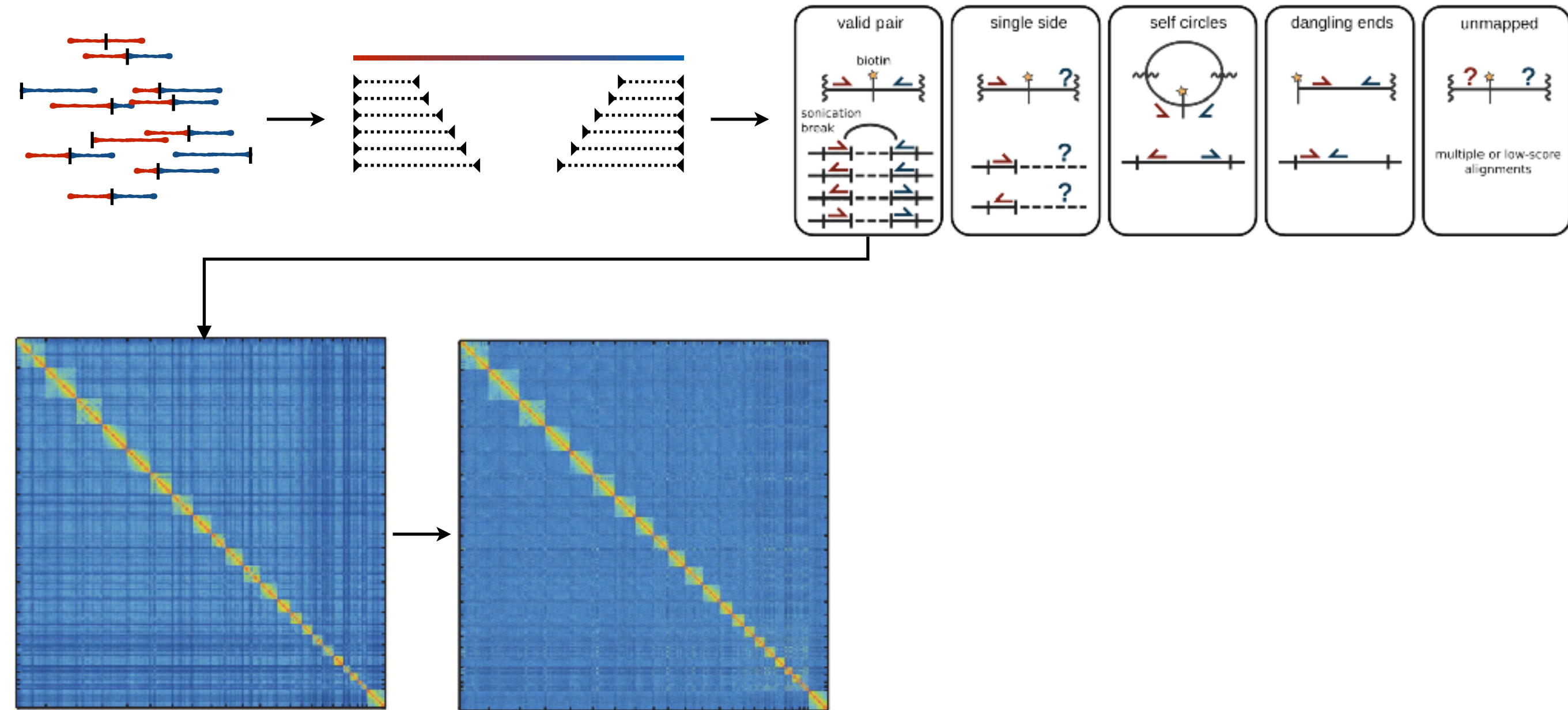
# From FASTQ to interaction matrices



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

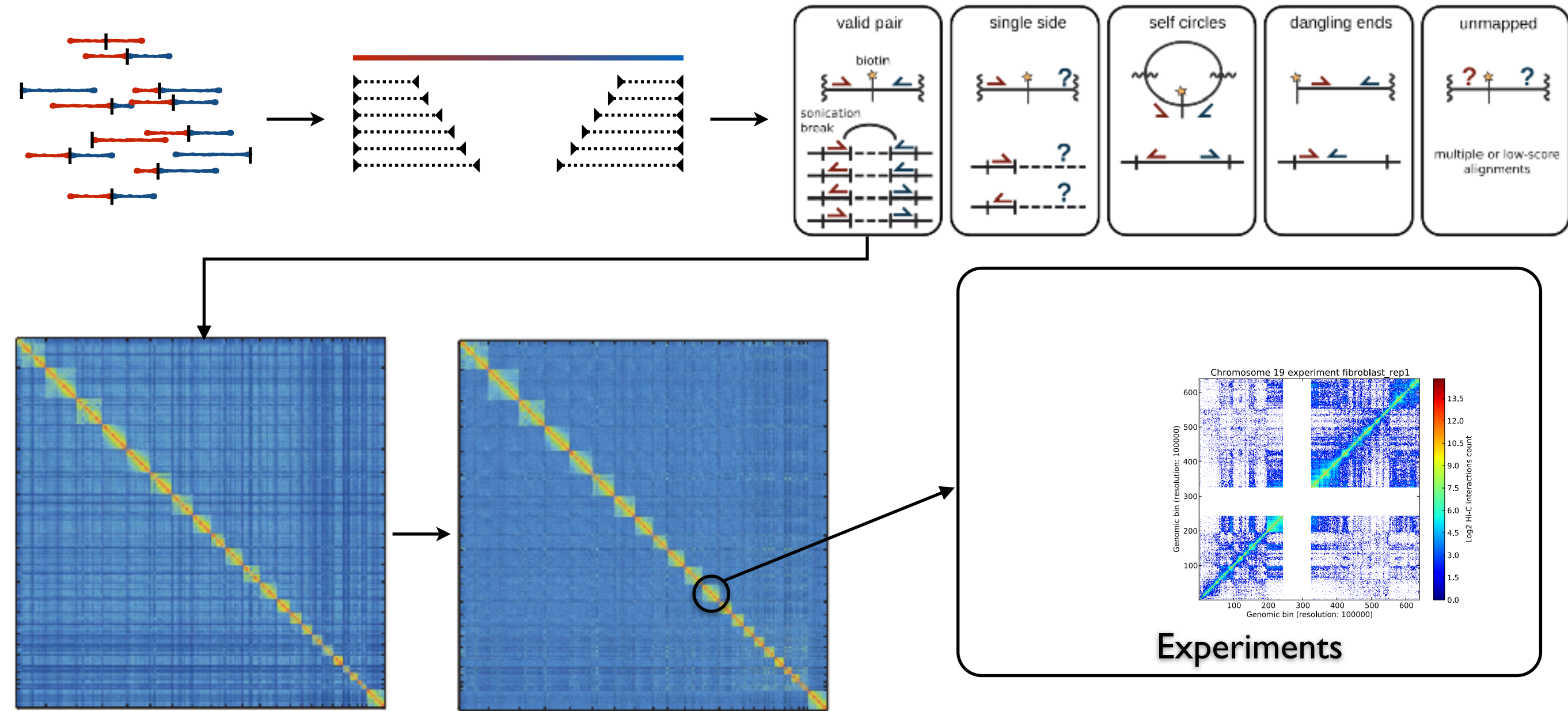
Iterative correction of Hi-C data reveals hallmarks of chromosome organization.  
Imakaev et al. Nature Methods (2012)

# From FASTQ to interaction matrices



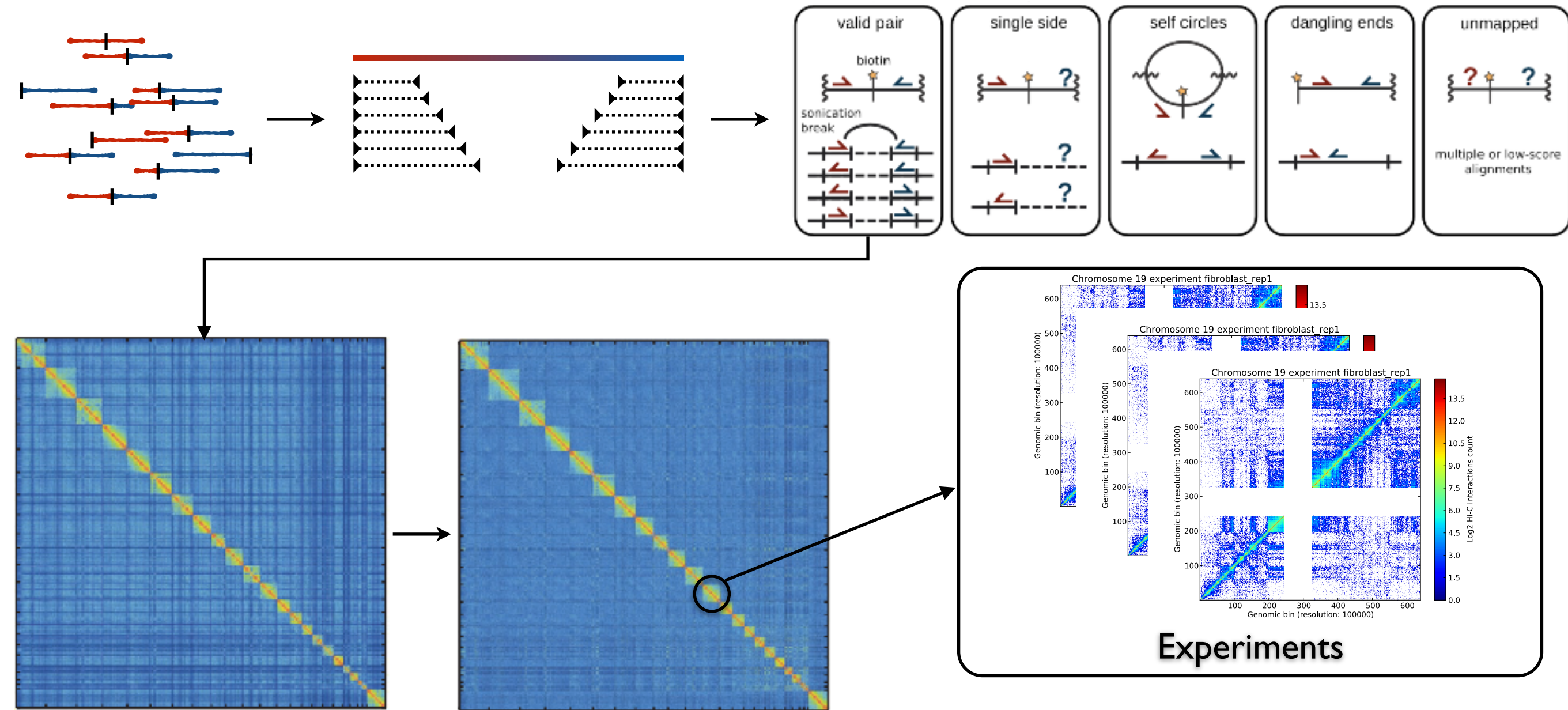


# From FASTQ to interaction matrices



Iterative correction of Hi-C data reveals hallmarks of chromosome organization.  
Imakaev et al. Nature Methods (2012)

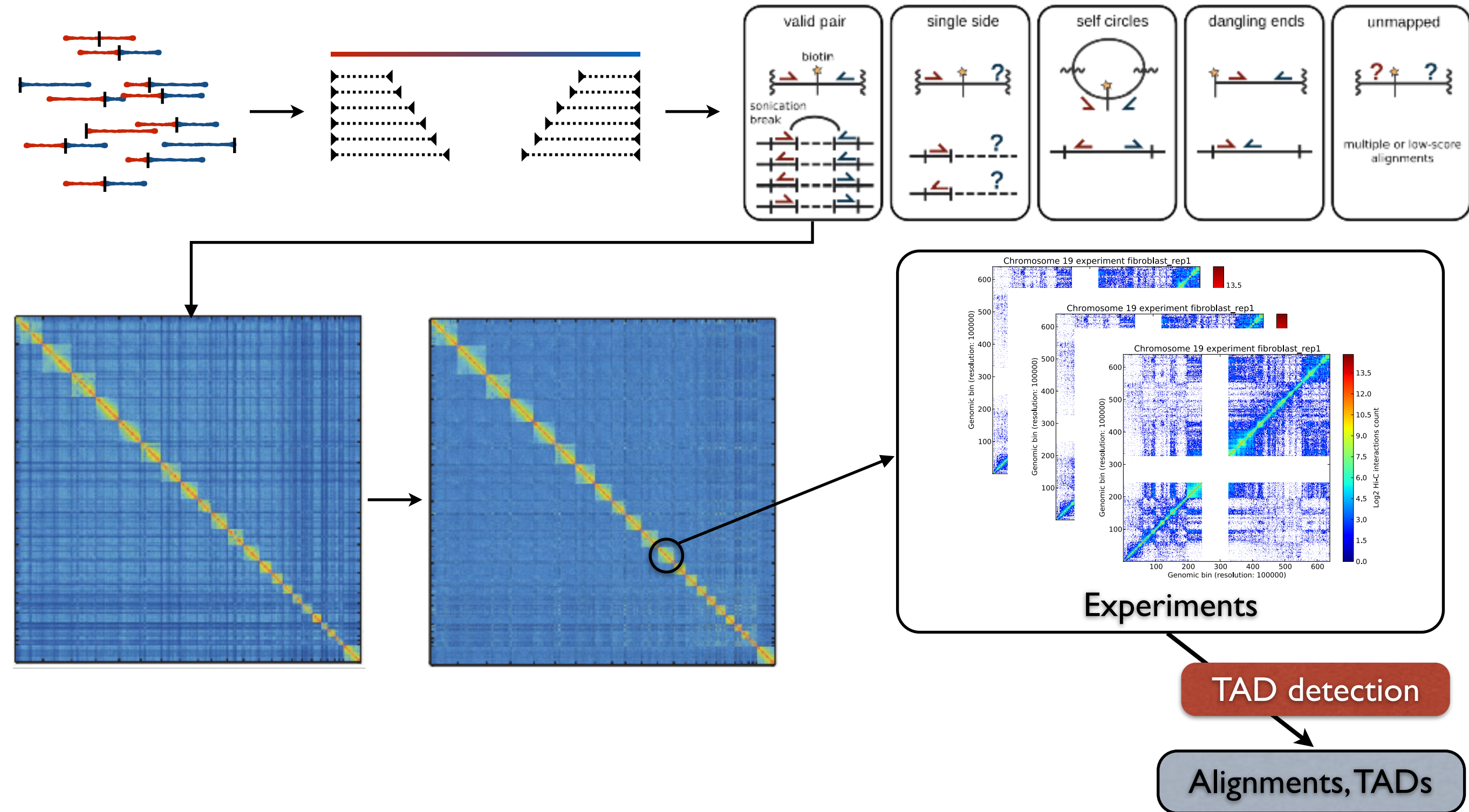
# From FASTQ to interaction matrices



Iterative correction of Hi-C data reveals hallmarks of chromosome organization.  
Imakaev et al. Nature Methods (2012)

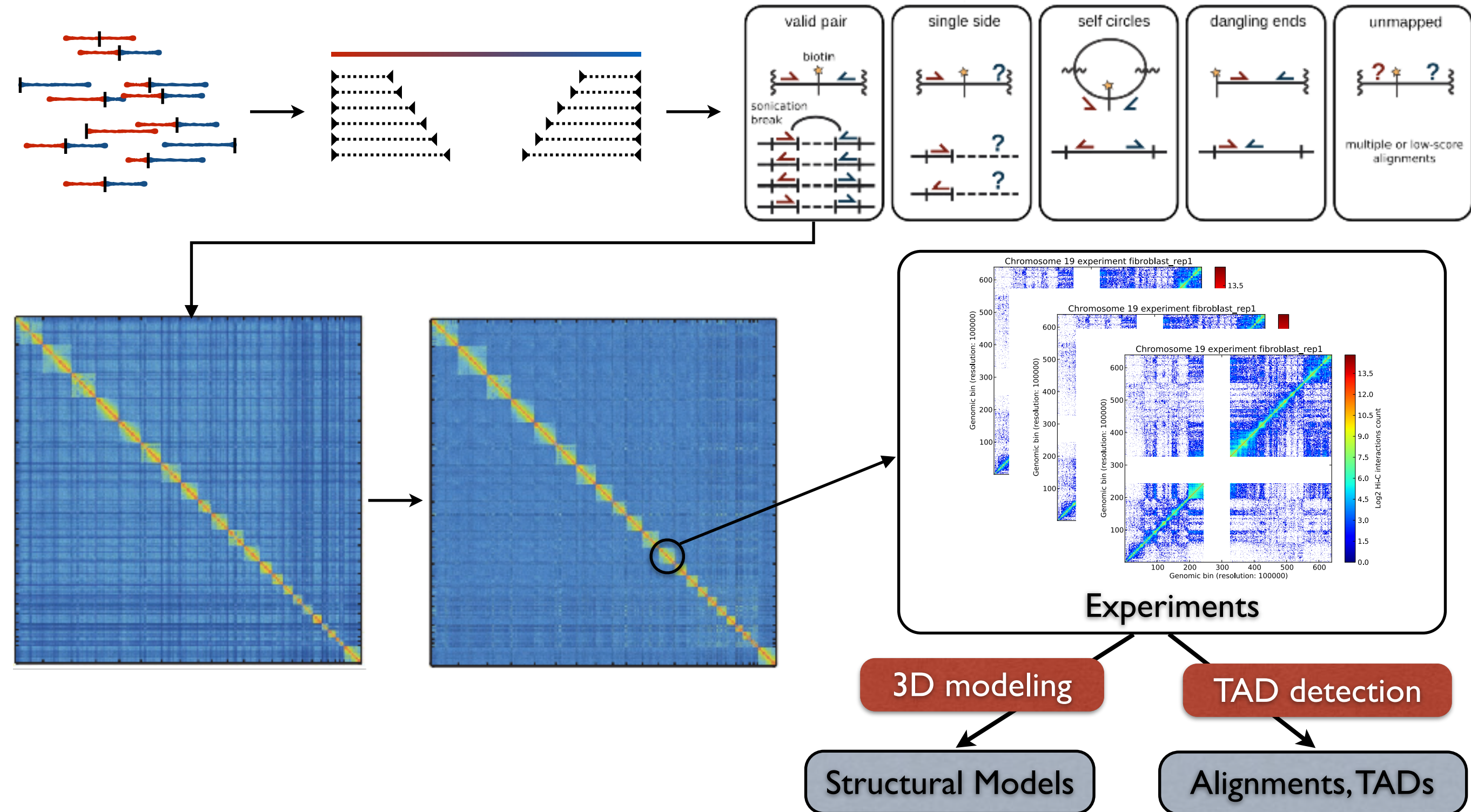


# From FASTQ to interaction matrices



Iterative correction of Hi-C data reveals hallmarks of chromosome organization.  
Imakaev et al. Nature Methods (2012)

# From FASTQ to interaction matrices



Iterative correction of Hi-C data reveals hallmarks of chromosome organization.  
Imakaev et al. Nature Methods (2012)



# Many alternatives

Tool	Short-read aligner(s)	Mapping improvement	Read filtering	Read-pair filtering	Normalization	Visualization	Confidence estimation	Implementation language(s)
HiCUP [46]	Bowtie/Bowtie2	Pre-truncation	✓	✓	—	—	—	Perl, R
Hiclib [47]	Bowtie2	Iterative	✓ <sup>a</sup>	✓	Matrix balancing	✓	—	Python
HiC-inspector [131]	Bowtie	—	✓	✓	—	✓	—	Perl, R
HIPPIE [132]	STAR	✓ <sup>b</sup>	✓	✓	—	—	—	Python, Perl, R
HiC-Box [133]	Bowtie2	—	✓	✓	Matrix balancing	✓	—	Python
HiCdat [122]	Subread	— <sup>c</sup>	✓	✓	Three options <sup>d</sup>	✓	—	C++, R
HiC-Pro [134]	Bowtie2	Trimming	✓	✓	Matrix balancing	—	—	Python, R
TADbit [120]	GEM	Iterative	✓	✓	Matrix balancing	✓	—	Python
HiOMER [62]	—	—	✓	✓	Two options <sup>e</sup>	✓	✓	Perl, R, Java
Hicpipe [54]	—	—	—	—	Explicit-factor	—	—	Perl, R, C++
HiBrowse [69]	—	—	—	—	—	✓	✓	Web-based
Hi-Corrector [57]	—	—	—	—	Matrix balancing	—	—	ANSI C
GOTHIC [135]	—	—	✓	✓	—	—	✓	R
HITC [121]	—	—	—	—	Two options <sup>f</sup>	✓	✓	R
chromoR [59]	—	—	—	—	Variance stabilization	—	—	R
HiFive [136]	—	—	✓	✓	Three options <sup>g</sup>	✓	—	Python
Fit-Hi-C [20]	—	—	—	—	—	✓	✓	Python

# Many alternatives

Tool	Short-read aligner(s)	Mapping improvement	Read filtering	Read-pair filtering	Normalization	Visualization	Confidence estimation	Implementation language(s)
HiCUP [46]	Bowtie/Bowtie2	Pre-truncation	✓	✓	—	—	—	Perl, R
Hiclib [47]	Bowtie2	Iterative	✓ <sup>a</sup>	✓	Matrix balancing	✓	—	Python
HiC-inspector [131]	Bowtie	—	✓	✓	—	✓	—	Perl, R
HIPPIE [132]	STAR	✓ <sup>b</sup>	✓	✓	—	—	—	Python, Perl, R
HiC-Box [133]	Bowtie2	—	✓	✓	Matrix balancing	✓	—	Python
HiCdat [122]	Subread	— <sup>c</sup>	✓	✓	Three options <sup>d</sup>	✓	—	C++, R
HiC-Pro [134]	Bowtie2	Trimming	✓	✓	Matrix balancing	—	—	Python, R
TADbit [120]	GEM	Iterative	✓	✓	Matrix balancing	✓	—	Python
HOMER [62]	—	—	✓	✓	Two options <sup>e</sup>	✓	✓	Perl, R, Java
Hicpipe [54]	—	—	—	—	Explicit-factor	—	—	Perl, R, C++
HiBrowse [69]	—	—	—	—	—	✓	✓	Web-based
Hi-Corrector [57]	—	—	—	—	Matrix balancing	—	—	ANSI C
GOTHIC [135]	—	—	✓	✓	—	—	✓	R
HITC [121]	—	—	—	—	Two options <sup>f</sup>	✓	✓	R
chromoR [59]	—	—	—	—	Variance stabilization	—	—	R
HiFive [136]	—	—	✓	✓	Three options <sup>g</sup>	✓	—	Python
Fit-Hi-C [20]	—	—	—	—	—	✓	✓	Python



# Many alternatives

Method <sup>*available online</sup>	Representation	Scoring				Sampling	Models
		$U_{3C}$		$U_{Biol}$	$U_{Phys}$		
		$F_{ij} \rightarrow D_{ij}$ conversion	Functional form				
ChromSDE* [37]	Points	$D_{ij} = \begin{cases} (F_{ij}^{-1})^\alpha & \text{if } F_{ij} > 0 \\ \infty & \text{if } F_{ij} = 0 \end{cases}$ $\alpha$ is optimized	$\sum_{(i,j) D_{ij}<\infty} \frac{(r_{ij}^2 - D_{ij}^2)}{D_{ij}} - \lambda \sum_{(i,j)} r_{ij}^2$ where $\lambda$ is set to 0.01	N/A	N/A	Deterministic semidefinite programming to find the coordinates	Consensus
ShRec3D* [38]	Points	$D_{ij} = \begin{cases} \left(\frac{1}{F_{ij}}\right)^\alpha & \text{if } F_{ij}' > 0 \\ \frac{N^2}{\sum_{k \neq i,j} F_{ik} F_{jk}} & \text{if } F_{ij}' = 0 \end{cases}$ $F_{ij}'$ is the original $F_{ij}$ corrected to satisfy all triangular inequalities with the shortest path reconstruction	N/A	N/A	N/A	Deterministic transformations of $D_{ij}$ into coordinates	Consensus
TADbit* [43]	Spheres	$D_{ij} \propto \begin{cases} \alpha F_{ij} + \beta & \text{if } F_{ij} < \gamma' \text{ or } F_{ij} > \gamma \\ \frac{s_i + s_j}{2} & \text{if }  i - j  = 1 \end{cases}$ $\alpha$ and $\beta$ are estimated from the max and the min $F_{ij}$ , from the optimized max distance and from the resolution. $\gamma' < \gamma$ are optimized too. $s_i$ is the radius of particle $i$	$\sum_{(i,j)} k_{ij} (r_{ij} - D_{ij})^2$ where $k_{ij} = 5$ if $ i - j  = 1$ or proportional to $F_{ij}$ otherwise	Yes	$U_{excl}$ and $U_{bond}$ have harmonic forms	Monte Carlo (MC) sampling with Simulated annealing and Metropolis scheme	Resampling
BACH* [45]	Points	$D_{ij} \propto \frac{B_i B_j}{F_{ij}^\alpha}$ . The biases $B_i$ and $B_j$ and $\alpha$ are optimized	$b_{ij} D_{ij}^{1/2} + c_{ij} \log(D_{ij})$ where $b_{ij}$ and $c_{ij}$ are optimized parameters	No	No	Sequential importance and Gibbs sampling with hybrid MC and adaptive rejection	Population
Giorgetti et al. [40]	Spheres	Particles interact with pair-wise well potentials of depths $B_{ij}$ and contact radius $a$ , which is larger than a hard-core radius and smaller than a maximum contact radius. The parameters are optimized over all the population of models		No	N/A	MC sampling with metropolis scheme	Population
Duan et al. [41]	Spheres	$\overline{F_{ i-j }} = \frac{\sum_{k= i-j }^{N- i-j } F_{(k+ i-j , k+ i-j )}}{N- i-j }$ is the average of $F_{ij}$ at genomic distance $ i - j $ expressed in kb. $D_{ij} = \overline{F_{ i-j }} \times 7.7 \times  i - j $ assuming that $\alpha$ 1 kb maps onto 7.7 nm	$\sum_{(i,j)} (r_{ij} - D_{ij})^2$	Yes	$U_{excl}$ and $U_{bond}$ have harmonic forms	Interior-point gradient-based method	Resampling
MCMC5C* [49]	Points	$D_{ij} \propto \frac{1}{F_{ij}^\alpha}$ where $\alpha$ is optimized	$\sum_{(i,j)} (F_{ij} - r_{ij}^{-1/2})^2$	N/A	N/A	MC sampling with Markov chain based algorithm	Resampling
PASTIS* [47]	Points	$D_{ij} \propto \frac{1}{F_{ij}^\alpha}$ where $\alpha$ is optimized	$b_{ij} D_{ij}^{1/2} + c_{ij} \log(D_{ij})$ where $b_{ij}$ and $c_{ij}$ are optimized parameters	No	No	Interior point and isotonic regression algorithms	Resampling
Meluzzi and Arya [48]	Spheres	$\sum_{(i,j)} k_{ij} r_{ij}^2$ where $k_{ij}$ are adjusted such that the contact probabilities computed on the models match the $F_{ij}$		No	$U_{excl}$ is a pure repulsive LJ potential. $U_{bond}$ and $U_{bend}$ have harmonic forms	Brownian dynamics	Resampling
AutoChrom3D* [44]	Points	$D_{ij} \propto \begin{cases} \alpha F_{ij} + \beta & \text{if } F_{min} < F_{ij} < F_\gamma \\ \alpha' F_{ij} + \beta' & \text{if } F_\gamma < F_{ij} < F_{max} \end{cases}$ where $F_{min}$ ( $F_{max}$ ) are the min(max) of $F_{ij}$ . The parameters $(\alpha, \beta)$ , $(\alpha', \beta')$ and $F_\gamma$ are found using the nuclear size, the resolution and the decay of $F_{ij}$ with $ i - j $	$\sum_{(i,j)} \frac{(r_{ij} - D_{ij})^2}{D_{ij}^2}$	Yes	N/A	Non-linear constrained	Consensus
Kalhor et al. [14]	Spheres	$D_{ij} = R_{contact}$ to enforce the pair contact, if the normalized contact frequency $F_{ij}$ is higher than 0.25. Otherwise the contact is not enforced	$\sum_{models} \sum_{(i,j)} k_{ij} (r_{ij} - D_{ij})^2$ where $k_{ij}$ is different for pairs of particles, on different chromosomes, on the same chromosome, or connected	Yes	$U_{excl}$ and $U_{bond}$ have harmonic forms	Conjugate gradients sampling with Simulated annealing scheme	Population

\* These methods are publicly available.