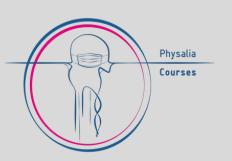
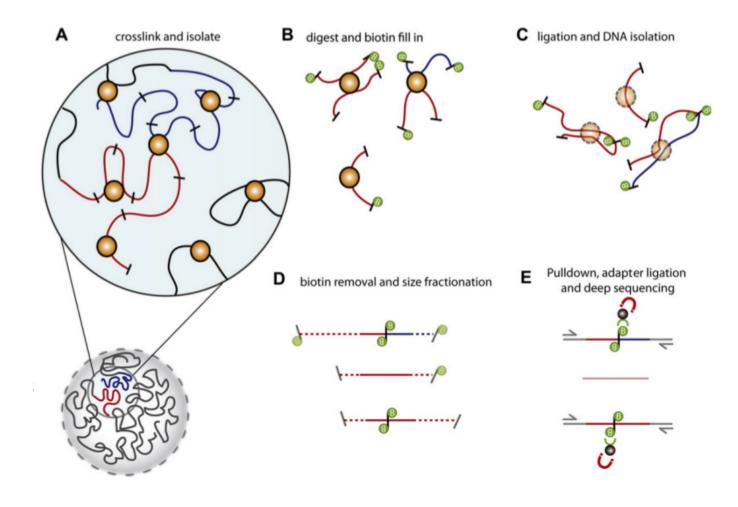
# Hi-C analysis

Epigenomics Data Analysis
Jacques Serizay
Physalia 2023

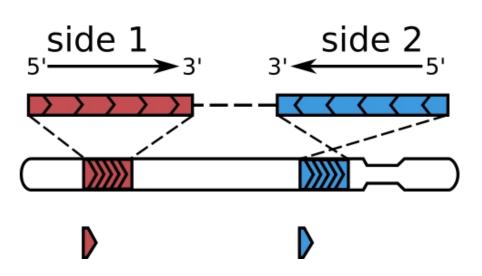


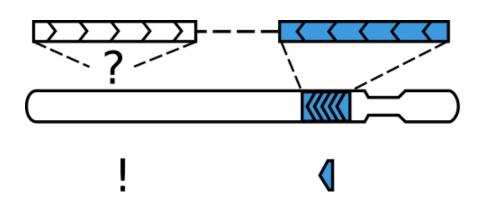
### **Hi-C** experimental workflow



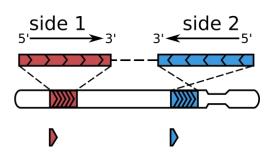
Belton et al., Methods 2012





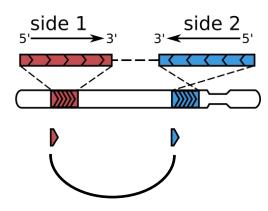






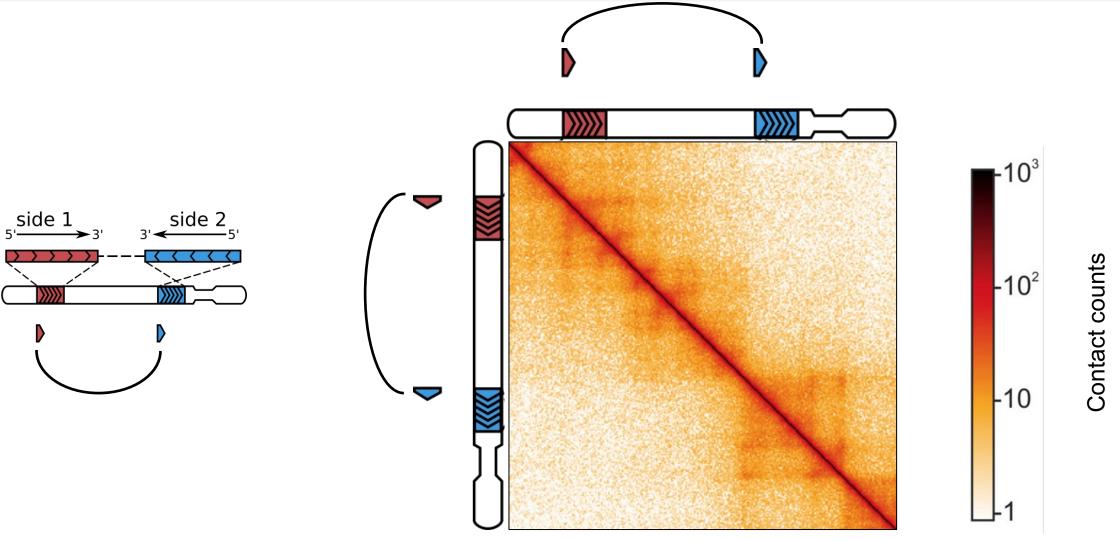
Open2C et al., Biorxiv 2023



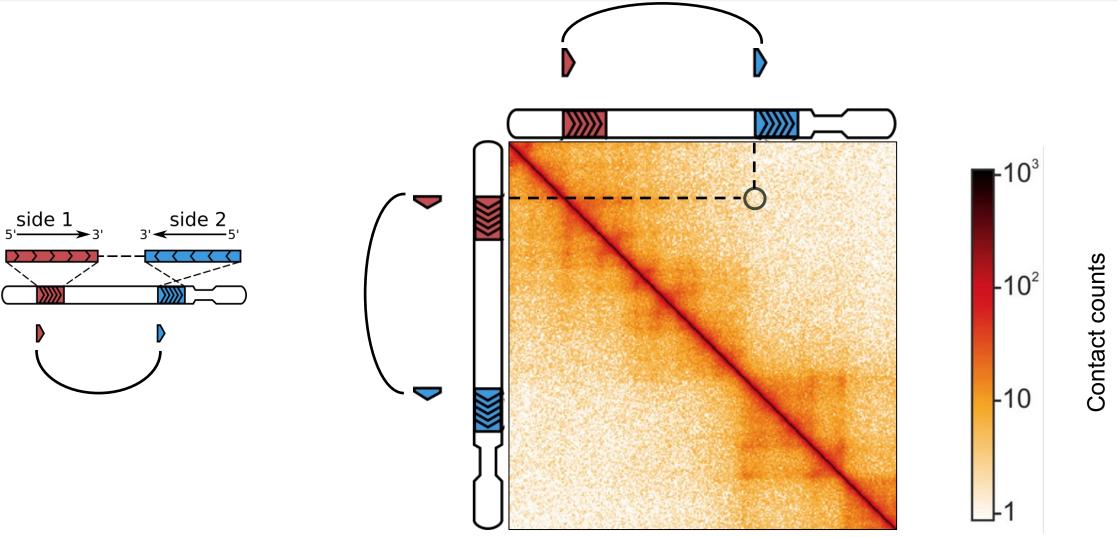


Open2C et al., Biorxiv 2023

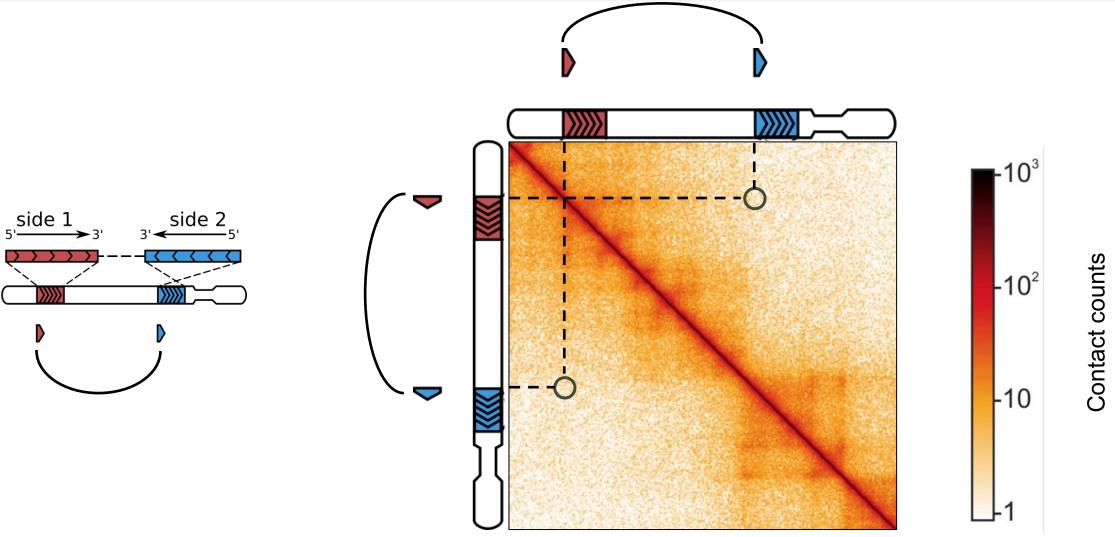






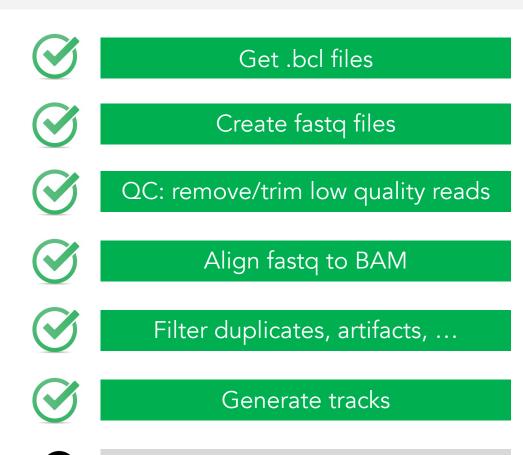








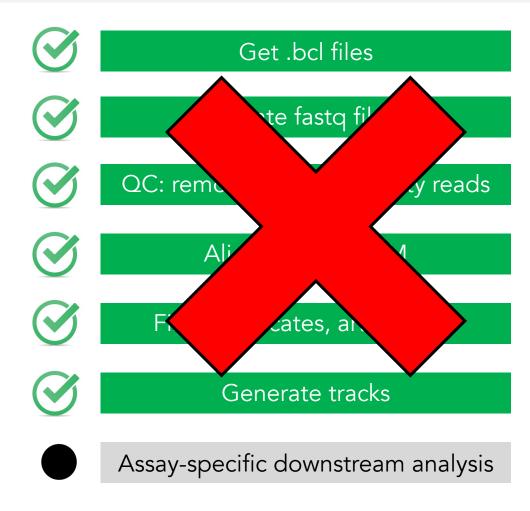
### **Hi-C pipeline**



Assay-specific downstream analysis

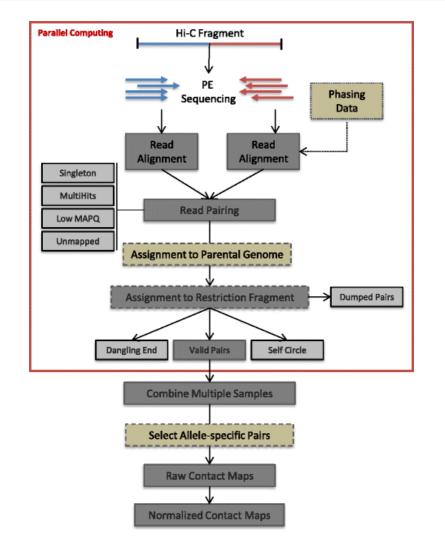


# **Hi-C pipeline**





### **Hi-C pipeline(s)**



FASTQ files Read mapping Use an aligner that Trim read end to 25bp Pre-truncate read ends allows split mappings containing the ligation junction sequence Post-process mapping results to filter out Map using a short-read aligner (Bowtie, Bowtie2, BWA, BWA-sw) ambiguous pairs SAM/BAM (b) Read level filtering \* 1.bam 🔻 \* 2.bam List of all restriction sites Keep only uniquely mapped and high quality (MAPQ ≥ 30) reads (digested genome) Assign read end to the Keep only reads proximal to the nearest site (MAPQ ≤ 300) nearest restriction site (binary search) SAM/BAM (C) Read-pair level filtering \*\_1.bam + Apply distance filters Apply strand filters Intra-chromasomal contacts below a distance threshold Туре Strand Valid pairs e.g. self ligation e.g. valid pair (discarded) <20-25 kb <20-25 kb Remove all but one Possible of duplicate pairs artifacts Valid pairs Normalization Bin using fixed-size windows (e.g., 40kb) or meta-fragments (e.g., 10 REs) Explicit-factor correction Matrix balancing Joint correction Corrects explicitly for known Assures uniform visibility for all Simultaneously corrects factors such as GC content. for explicit factors and loci (i.e. equal row sum). Corrects for known and genomic distance between fragment length and potentially unknown biases mappability interacting loci

11

Servant et al., Genome Biol. 2015 Ay & Noble Genome Biol 2015



### Hi-C pipeline(s) – 2015

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	$\checkmark^a$	$\checkmark$	Matrix balancing	✓	_	Python
HiC-inspector [131]	Bowtie	_	$\checkmark$	$\checkmark$	_	✓	_	Perl, R
HIPPIE [132]	STAR	√b	$\checkmark$	$\checkmark$	_	_	_	Python, Perl, R
HiC-Box [133]	Bowtie2	_	✓	$\checkmark$	Matrix balancing	✓	_	Python
HiCdat [122]	Subread	_c	✓	$\checkmark$	Three options <sup>d</sup>	$\checkmark$	_	C++, R
HiC-Pro [134]	Bowtie2	Trimming	✓	$\checkmark$	Matrix balancing	_	_	Python, R
TADbit [120]	GEM	Iterative	$\checkmark$	$\checkmark$	Matrix balancing	✓	_	Python
HOMER [62]	_	_	✓	$\checkmark$	Two options <sup>e</sup>	✓	$\checkmark$	Perl, R, Java
Hicpipe [54]	_	_	_	_	Explicit-factor	_	_	Perl, R, C++
HiBrowse [69]	_	_	_	_	_	✓	✓	Web-based
Hi-Corrector [57]	_	_	_	_	Matrix balancing	_	_	ANSI C
GOTHIC [135]	_	_	✓	$\checkmark$	_	_	$\checkmark$	R
HiTC [121]	_	_	_	_	Two options <sup>f</sup>	$\checkmark$	$\checkmark$	R
chromoR [59]	_	_	_	_	Variance stabilization	_	_	R
HiFive [136]	_	_	✓	$\checkmark$	Three options $^g$	✓	_	Python
Fit-Hi-C [20]	_	_	_	_	_	✓	$\checkmark$	Python

<sup>&</sup>lt;sup>a</sup>Hiclib keeps the reads with only one mapped end (single-sided reads) for use in coverage computations

Servant et al., Genome Biol. 2015 Ay & Noble Genome Biol 2015



<sup>&</sup>lt;sup>b</sup>HIPPIE states that it rescues chimeric reads. No details are given

<sup>&</sup>lt;sup>c</sup>HiCdat reports no substantial improvement in successfully aligned read pairs when iterative mapping in Hiclib is used for *Arabidopsis thaliana* Hi-C data

<sup>&</sup>lt;sup>d</sup>HiCdat provides three options for normalization: coverage and distance correction, HiCNorm and ICE

eHOMER provides two options for normalization: simpleNorm corrects for sequencing coverage only and norm corrects for coverage plus the genomic distance between loci fHiTC provides two options for normalization: normLGF implements HiCNorm and normICE implements ICE algorithm from Hiclib

<sup>&</sup>lt;sup>9</sup>HiFive provides three options - Probability, Express, and Binning - for normalization. The Express and Binning algorithms correspond to matrix balancing and explicit-factor correction schemes, respectively