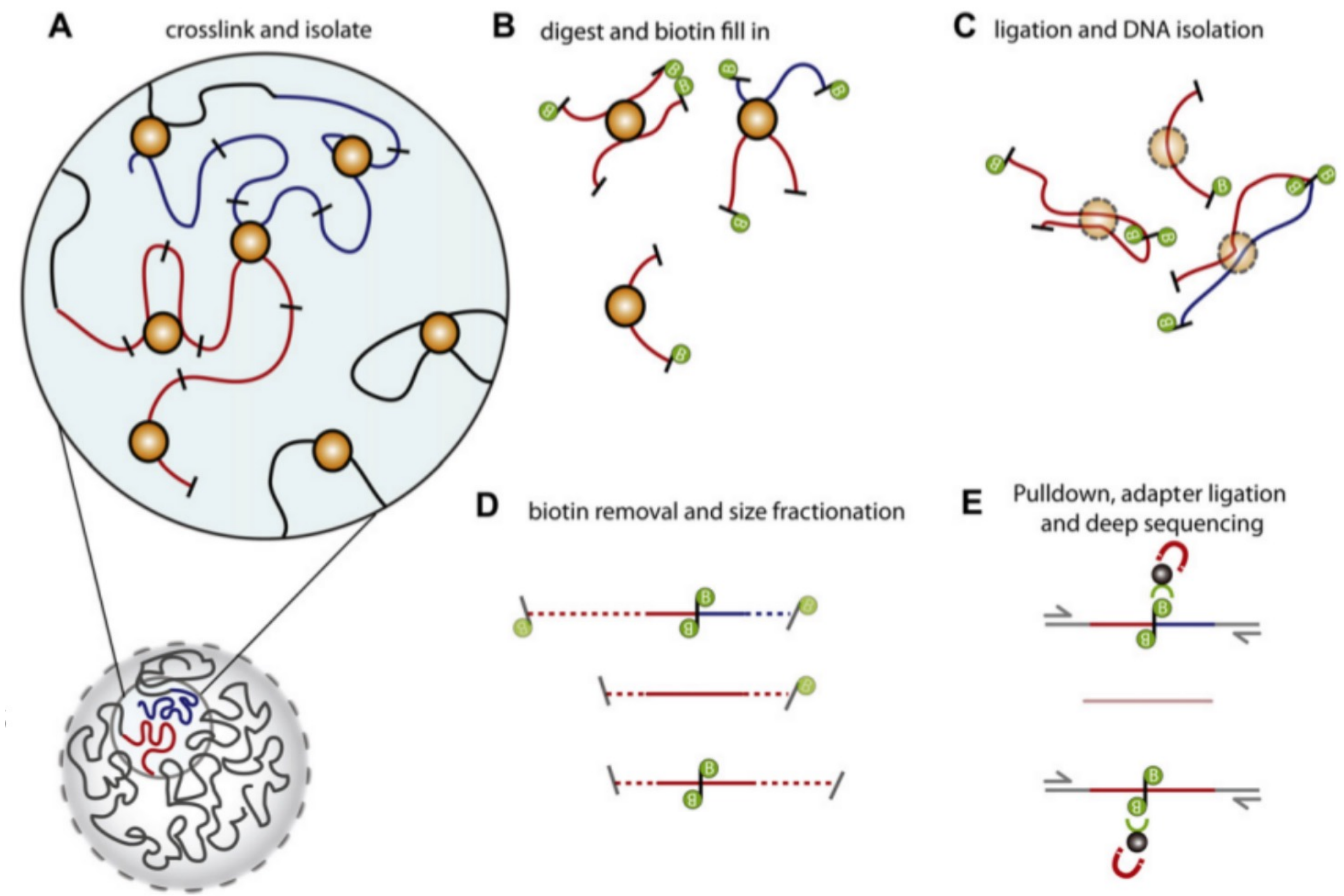
Hi-C analysis

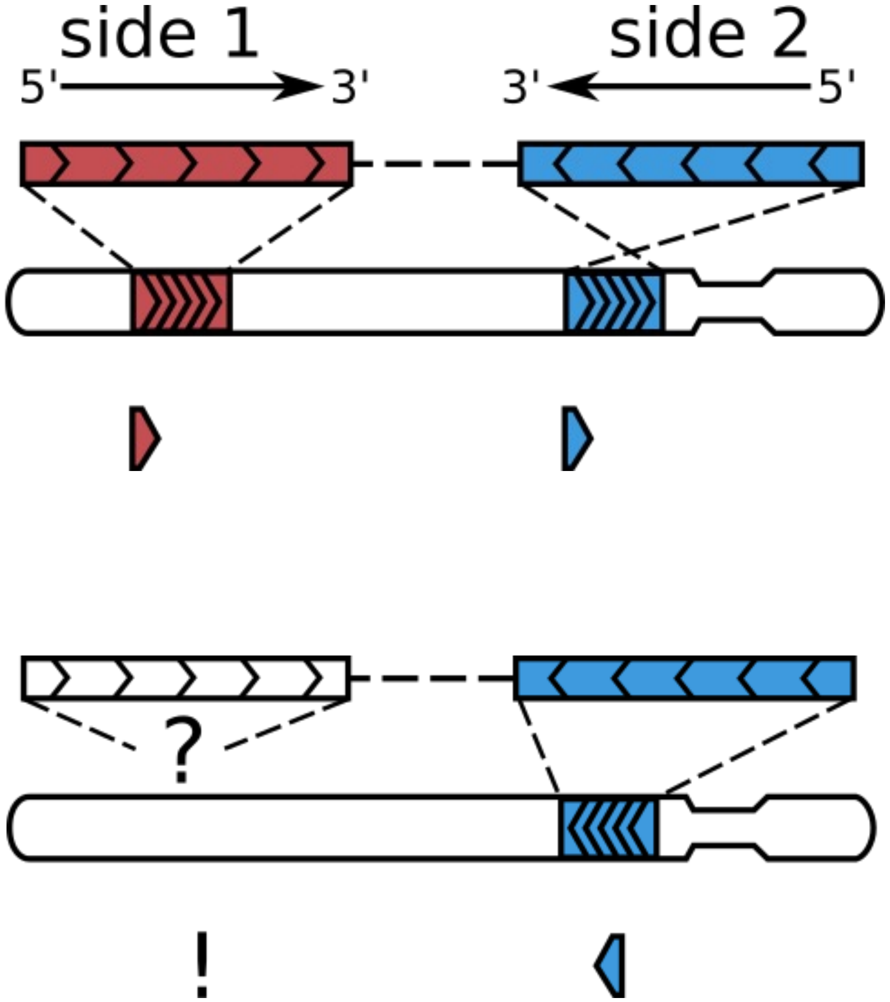
Epigenomics Data Analysis
Jacques Serizay
Physalia 2023

Hi-C experimental workflow



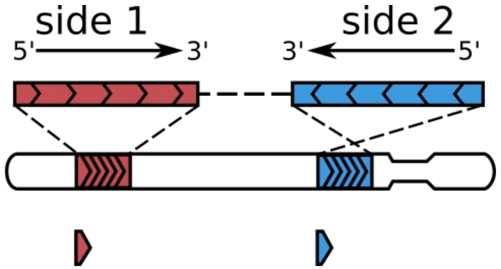
Belton et al., Methods 2012

Hi-C computational workflow



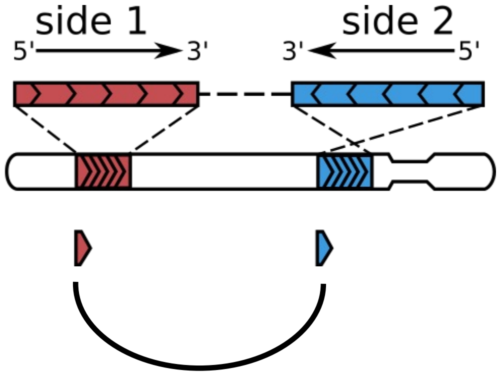
Open2C et al., Biorxiv 2023

Hi-C computational workflow



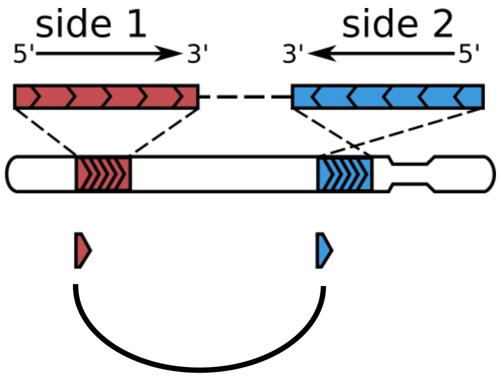
Open2C et al., Biorxiv 2023

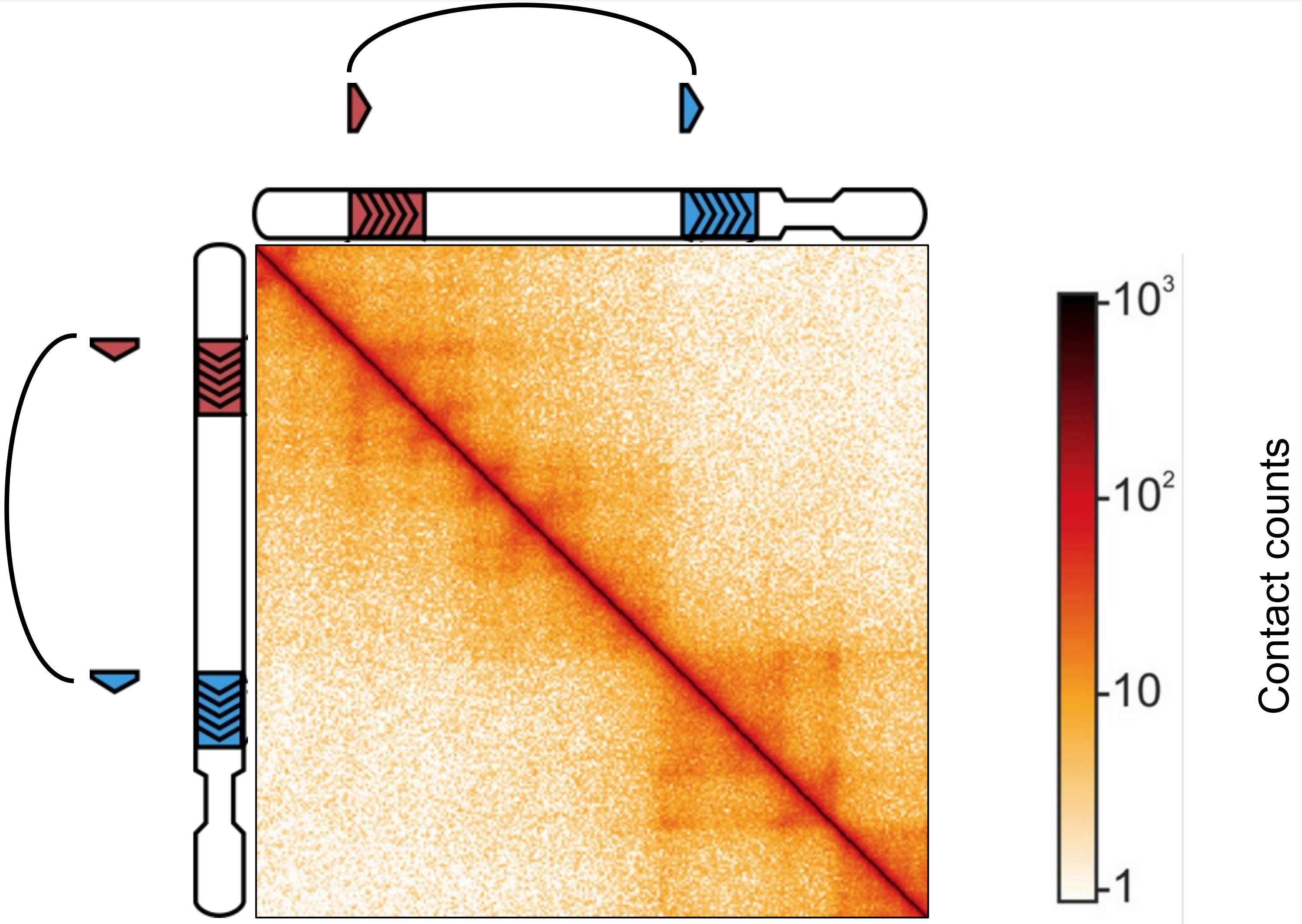
Hi-C computational workflow



Open2C et al., Biorxiv 2023

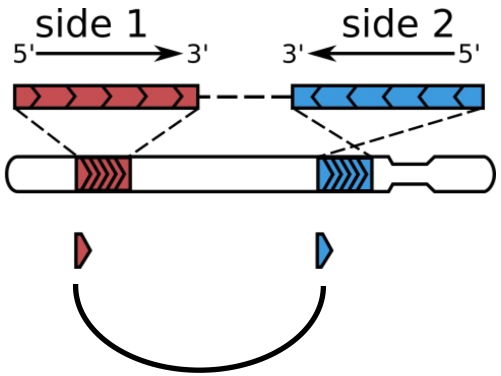
Hi-C computational workflow

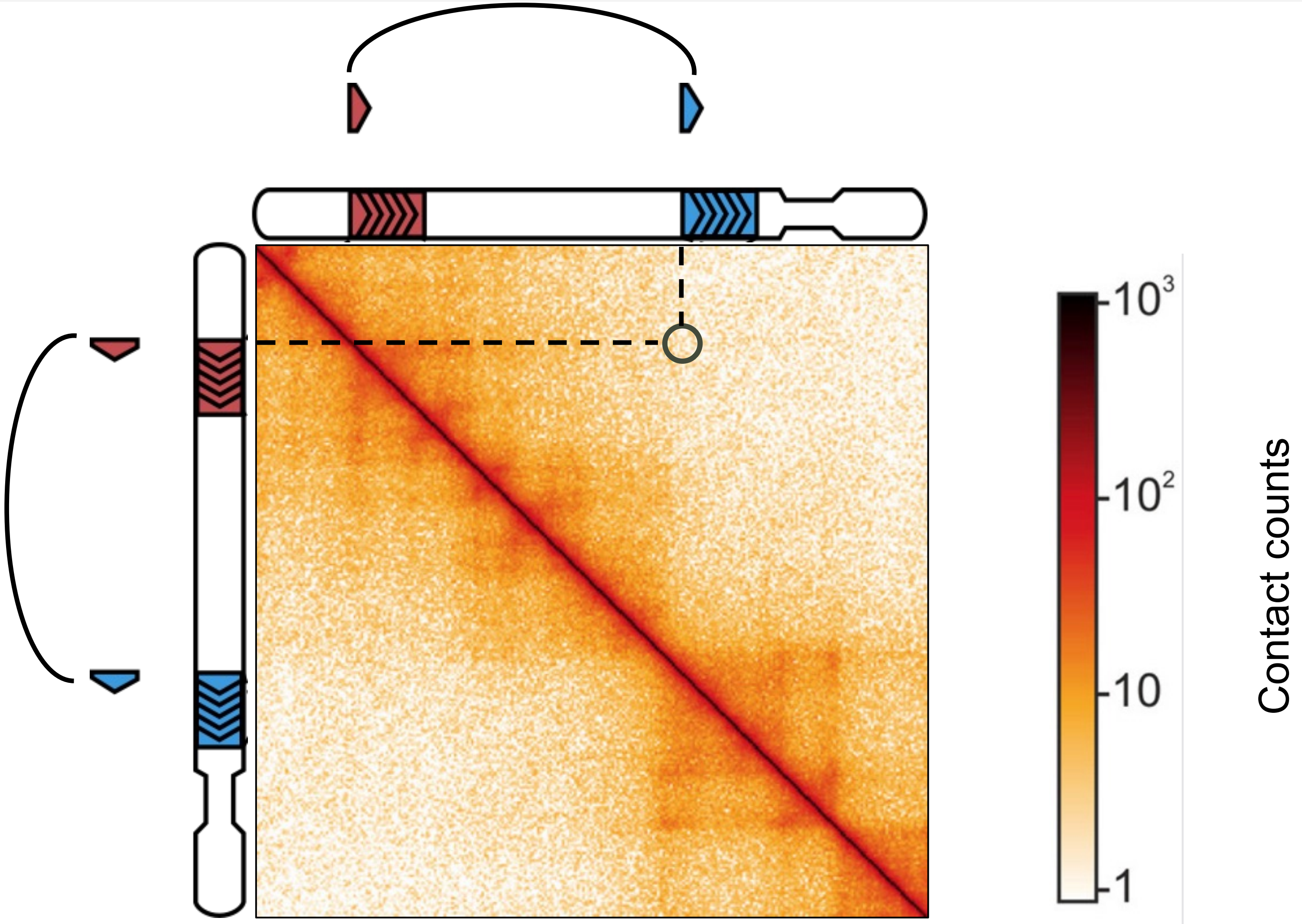




Open2C et al., Biorxiv 2023

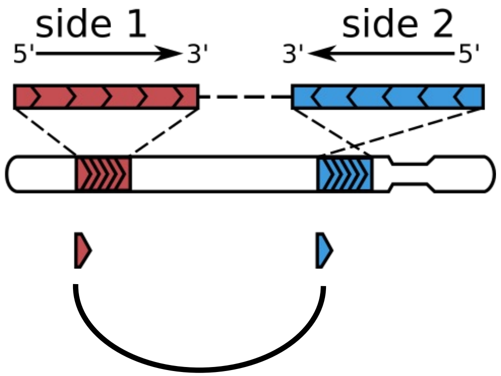
Hi-C computational workflow

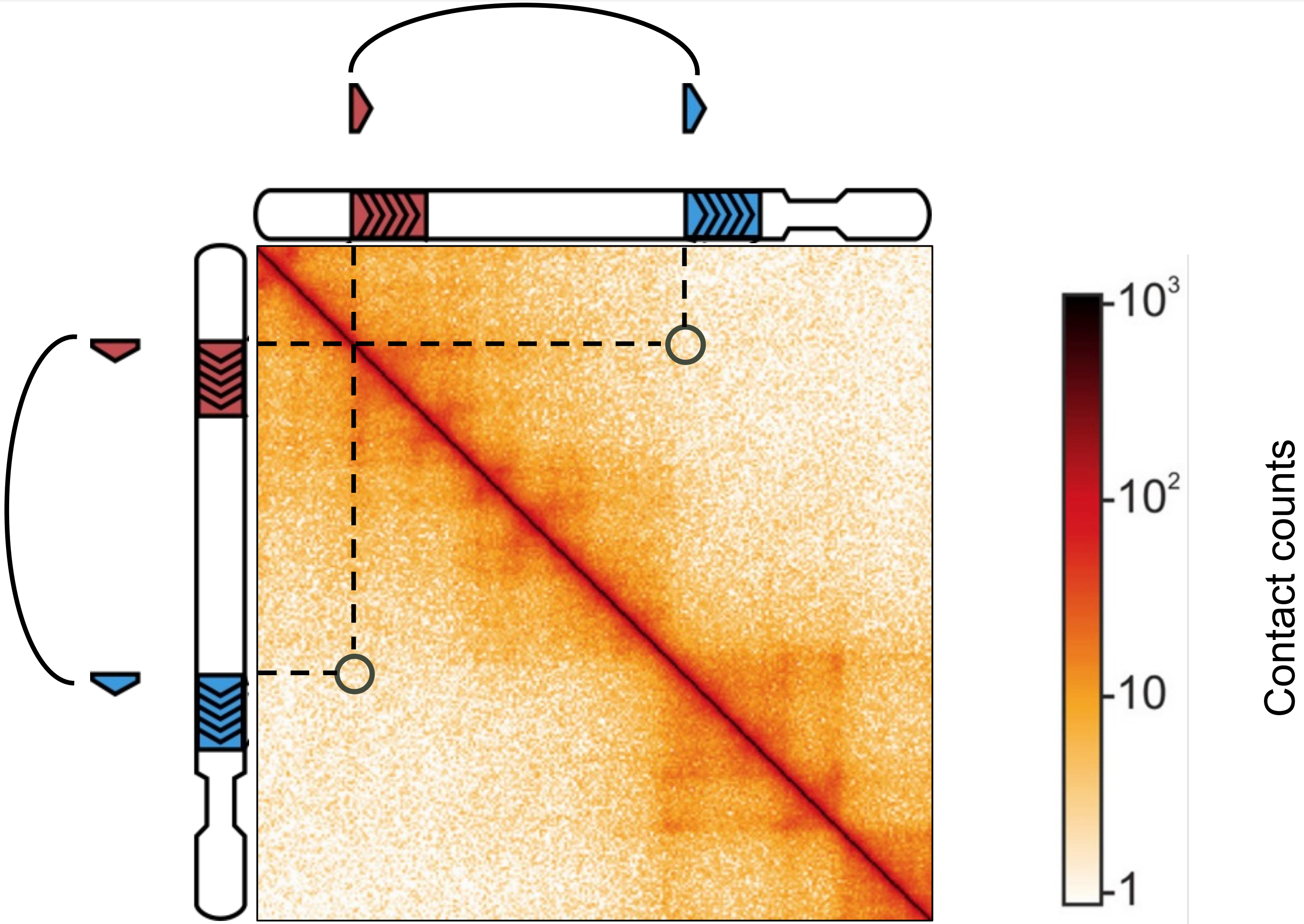




Open2C et al., Biorxiv 2023

Hi-C computational workflow





Open2C et al., Biorxiv 2023

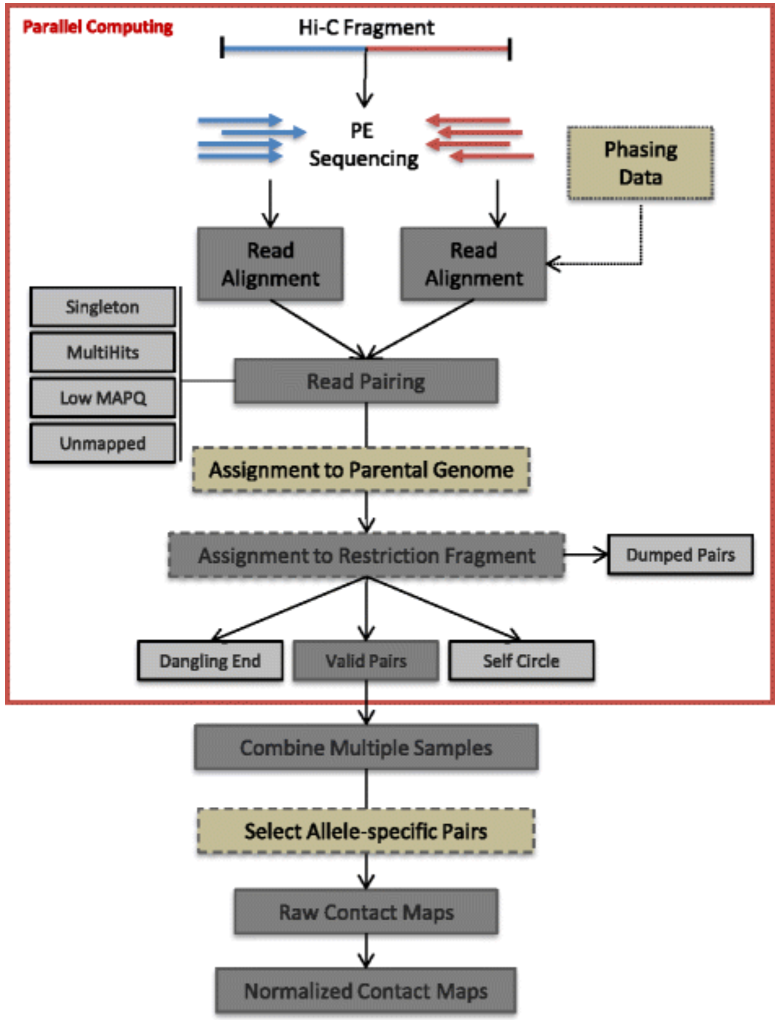
Hi-C pipeline

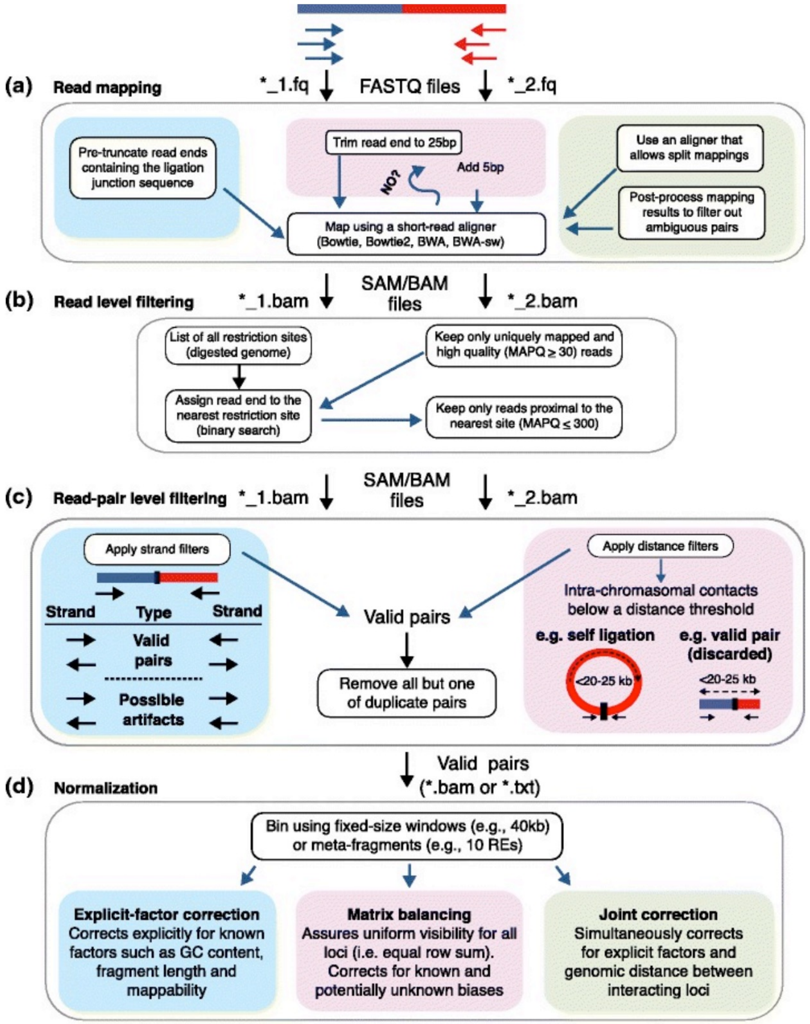
- ✓ Get .bcl files
- ✓ Create fastq files
- ✓ QC: remove/trim low quality reads
- ✓ Align fastq to BAM
- ✓ Filter duplicates, artifacts, …
- ✓ Generate tracks
- ● Assay-specific downstream analysis

Hi-C pipeline

- ✓ Get .bcl files
- ✓ te fastq fil
- ✓ QC: remo y reads
- ✓ Ali M
- ✓ Fi cates, ar
- ✓ Generate tracks
- ● Assay-specific downstream analysis

Hi-C pipeline(s)

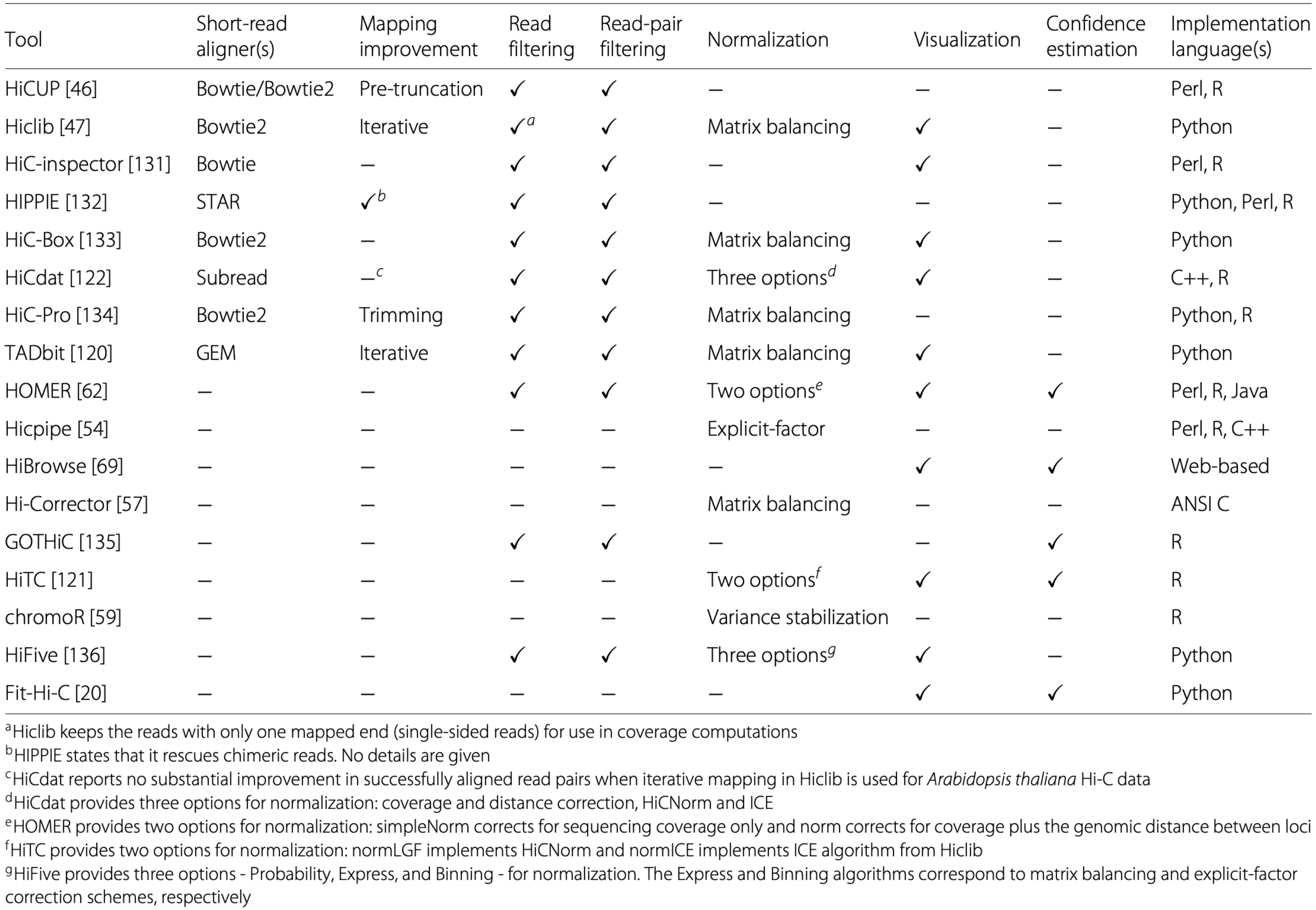




Servant et al., Genome Biol. 2015

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	✓[a]	✓	Matrix balancing	✓	–	Python
HiC-inspector [131]	Bowtie	–	✓	✓	–	✓	–	Perl, R
HIPPIE [132]	STAR	✓[b]	✓	✓	–	–	–	Python, Perl, R
HiC-Box [133]	Bowtie2	–	✓	✓	Matrix balancing	✓	–	Python
HiCdat [122]	Subread	–[c]	✓	✓	Three options[d]	✓	–	C++, R
HiC-Pro [134]	Bowtie2	Trimming	✓	✓	Matrix balancing	–	–	Python, R
TADbit [120]	GEM	Iterative	✓	✓	Matrix balancing	✓	–	Python
HOMER [62]	–	–	✓	✓	Two options[e]	✓	✓	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[f]	✓	✓	R
chromoR [59]	–	–	–	–	Variance stabilization	–	–	R
HiFive [136]	–	–	✓	✓	Three options[g]	✓	–	Python
Fit-Hi-C [20]	–	–	–	–	–	✓	✓	Python

[a] Hiclib keeps the reads with only one mapped end (single-sided reads) for use in coverage computations

[b] HIPPIE states that it rescues chimeric reads. No details are given

[c] HiCdat reports no substantial improvement in successfully aligned read pairs when iterative mapping in Hiclib is used for *Arabidopsis thaliana* Hi-C data

[d] HiCdat provides three options for normalization: coverage and distance correction, HiCNorm and ICE

[e] HOMER provides two options for normalization: simpleNorm corrects for sequencing coverage only and norm corrects for coverage plus the genomic distance between loci

[f] HiTC provides two options for normalization: normLGF implements HiCNorm and normICE implements ICE algorithm from Hiclib

[g] 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

Servant et al., Genome Biol. 2015

Ay & Noble Genome Biol 2015