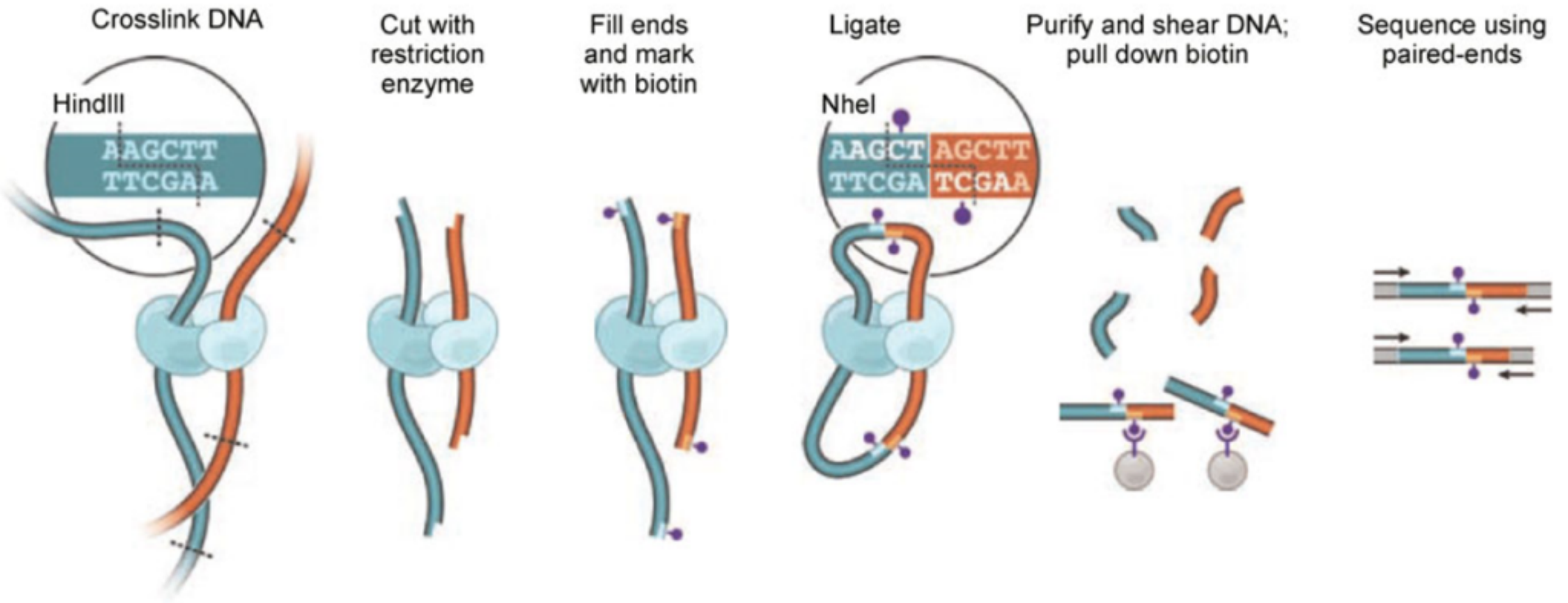
What we need to know about NGS in Hi-C

François Serra, David Castillo & Marc A. Marti-Renom
Structural Genomics Group (CNAG-CRG)

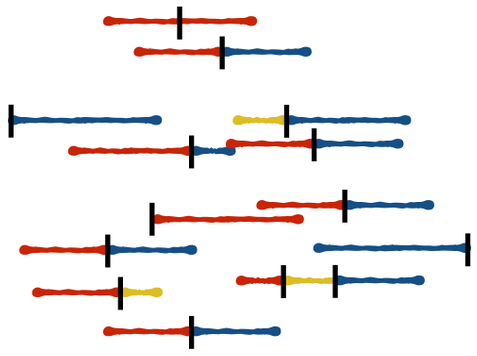
Illumina official video

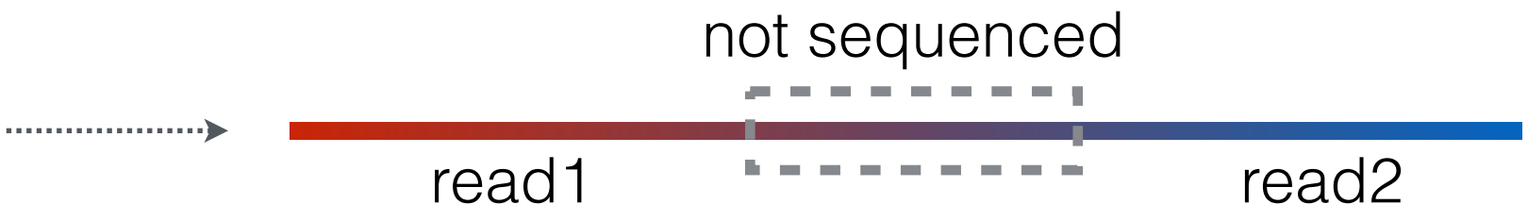
Hi-C experiment



Lieberman-Aiden, E. … Dekker, J. (2009). Science, 326(5950),

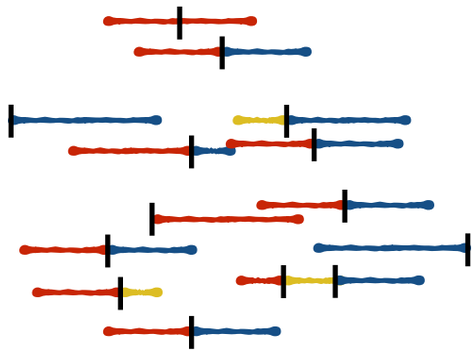
HiC mapping

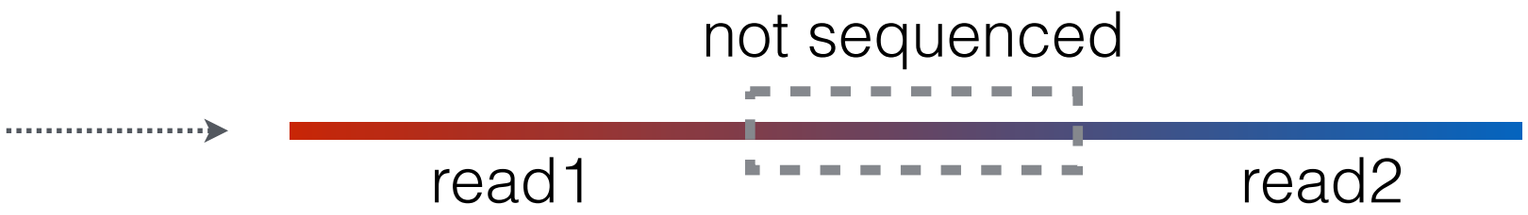




Iterative mapping

HiC mapping

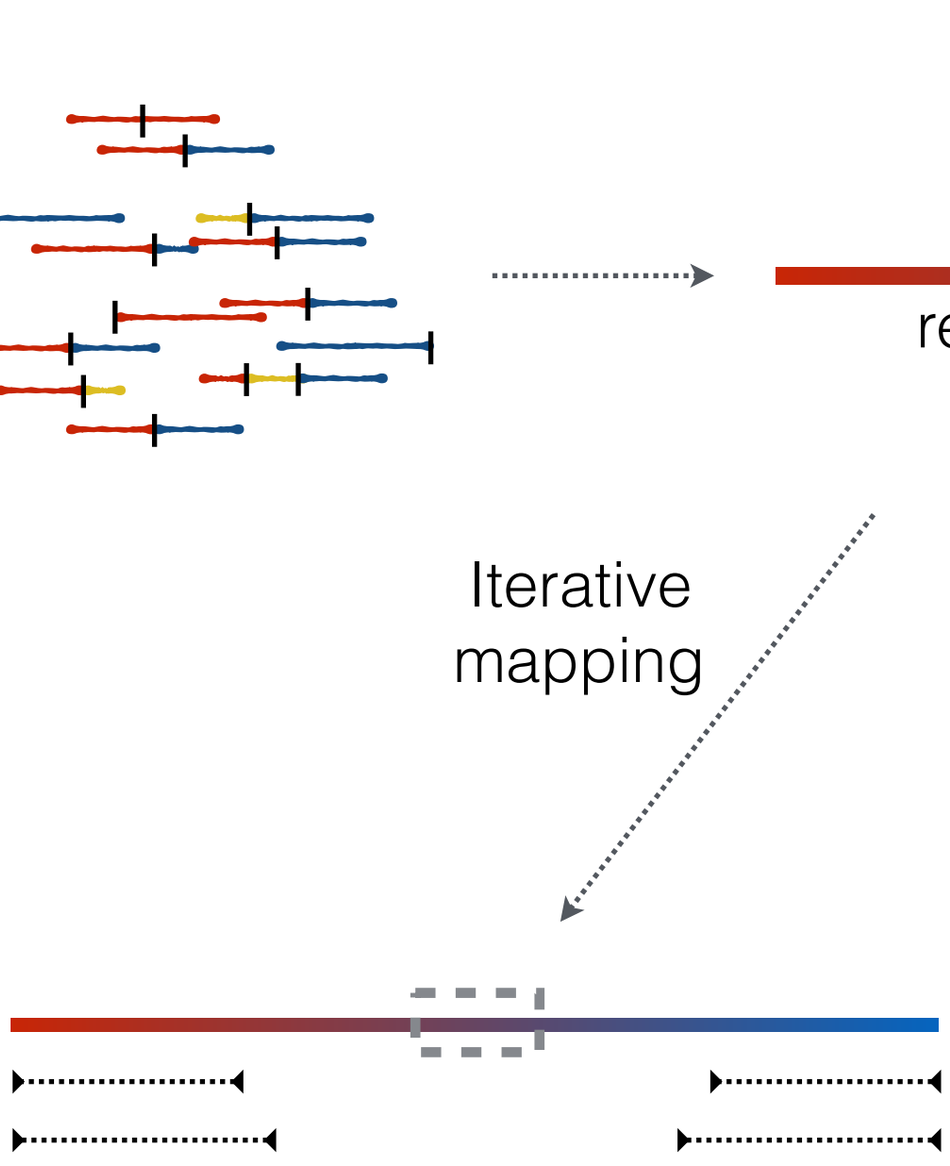




Iterative mapping

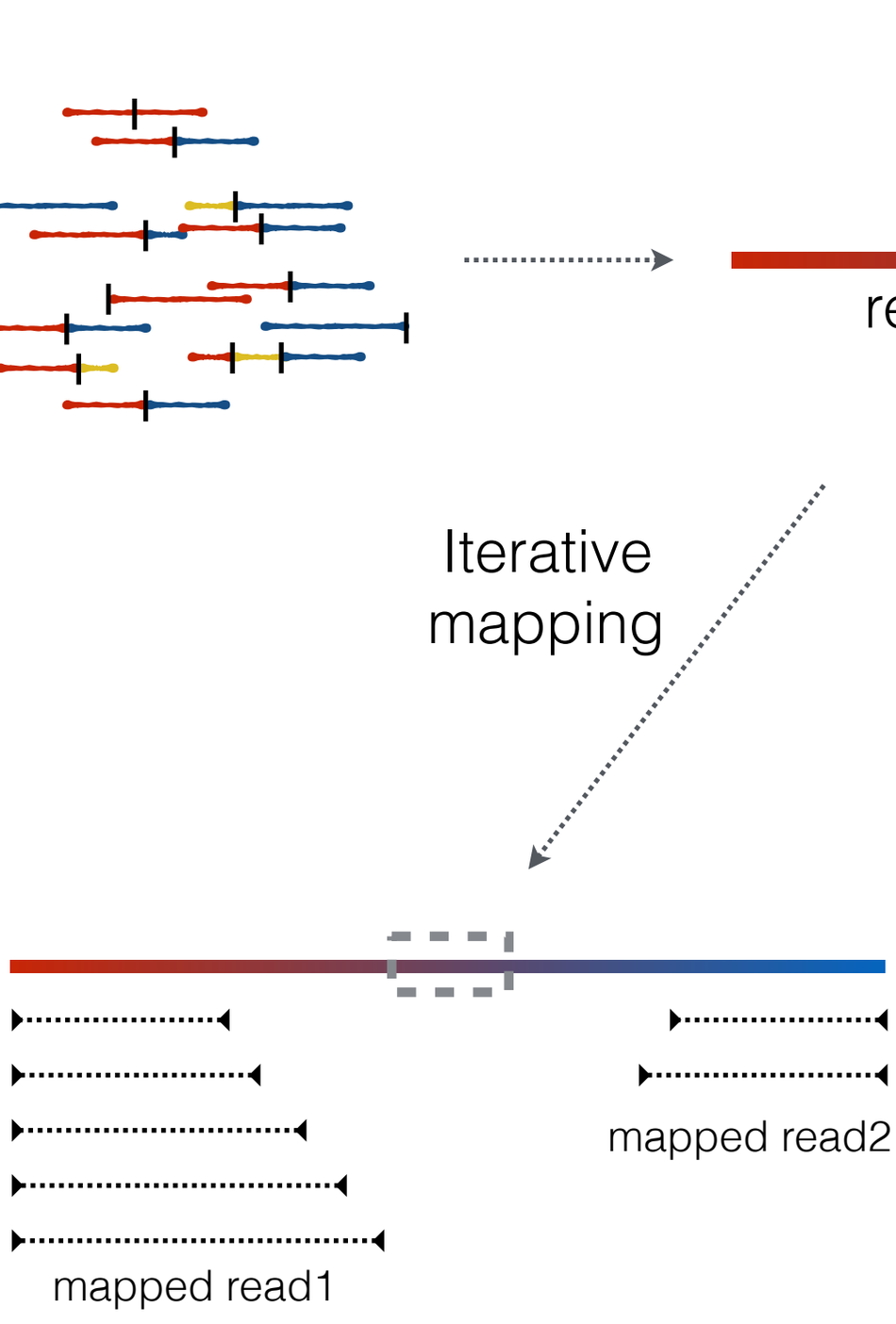
HiC mapping

Iterative mapping

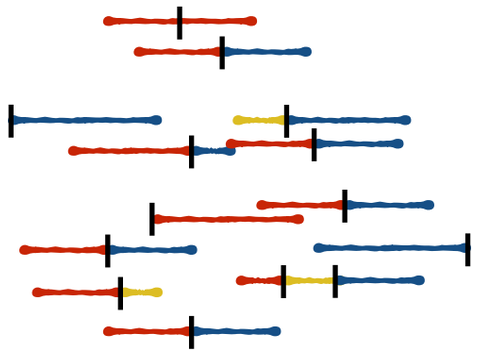


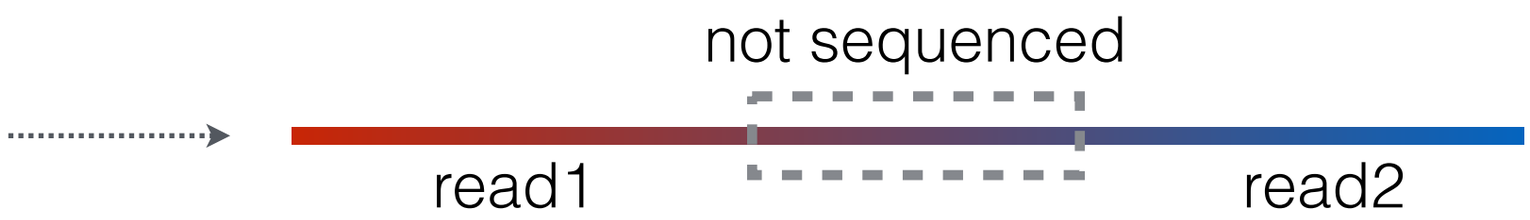
HiC mapping

Iterative mapping



HiC mapping

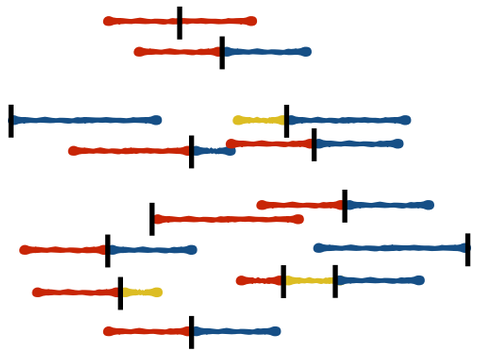


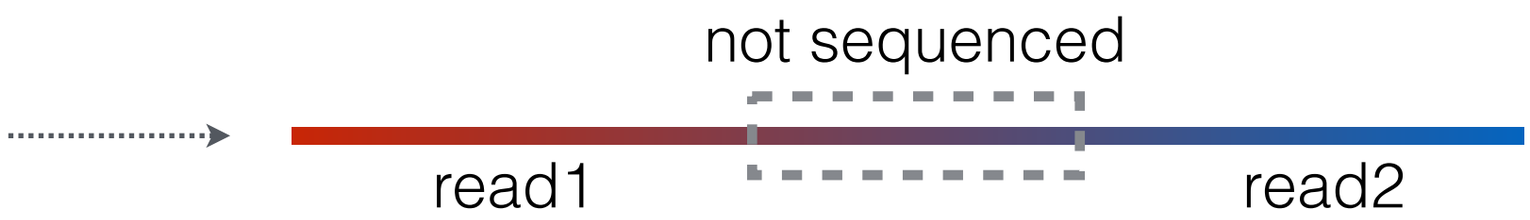


Iterative mapping

fragment based mapping

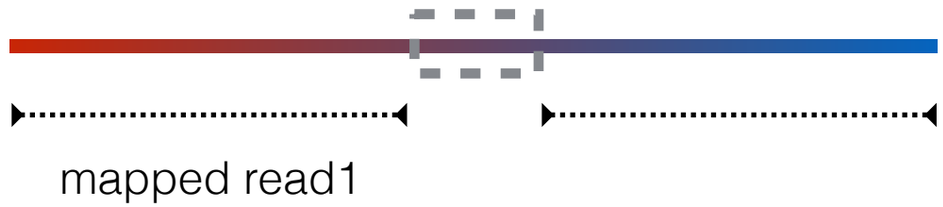
HiC mapping



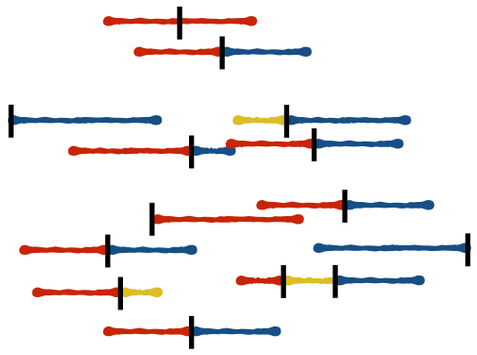


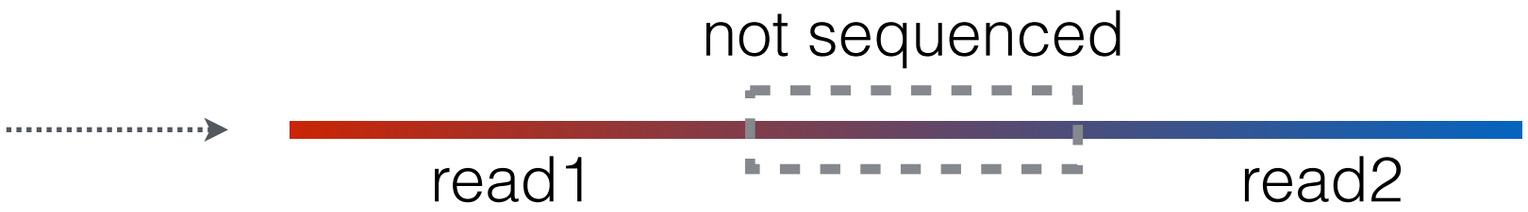
Iterative mapping

fragment based mapping

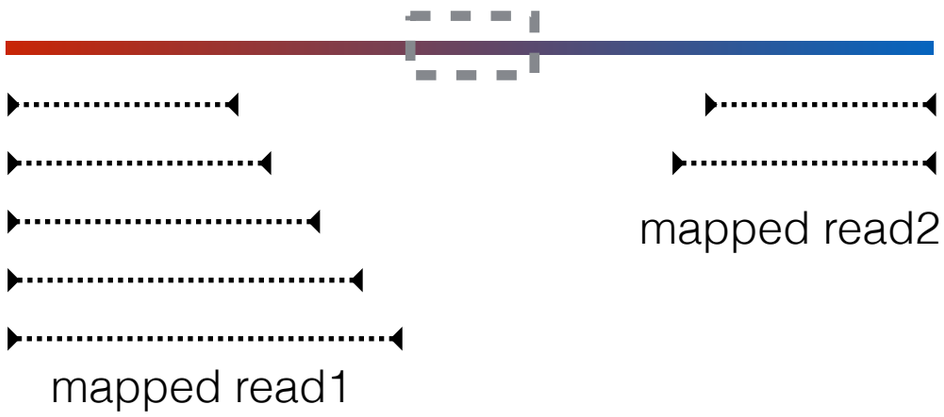


HiC mapping

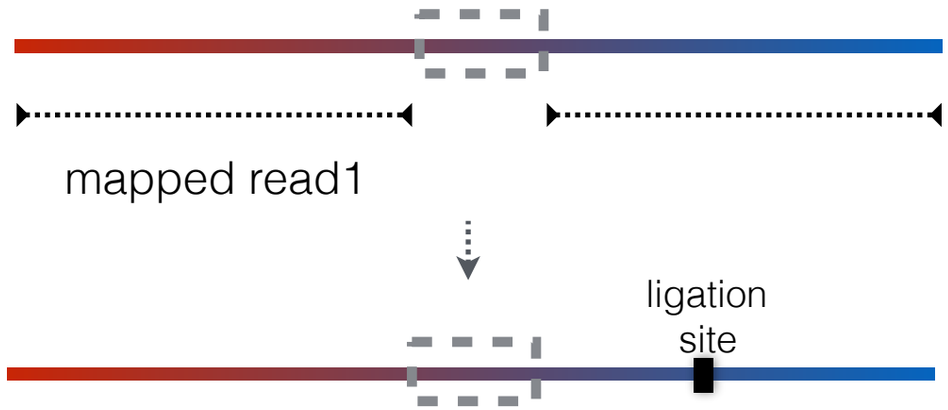




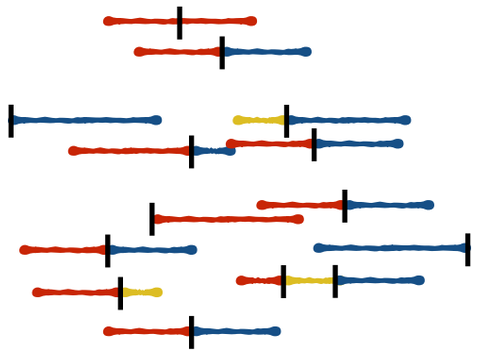
Iterative mapping

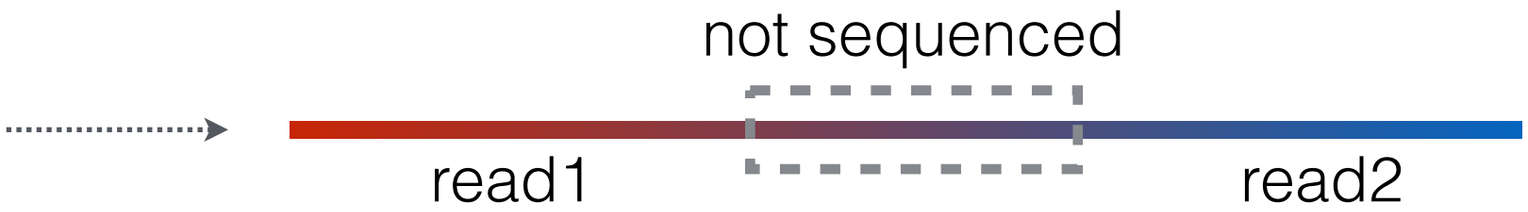


fragment based mapping



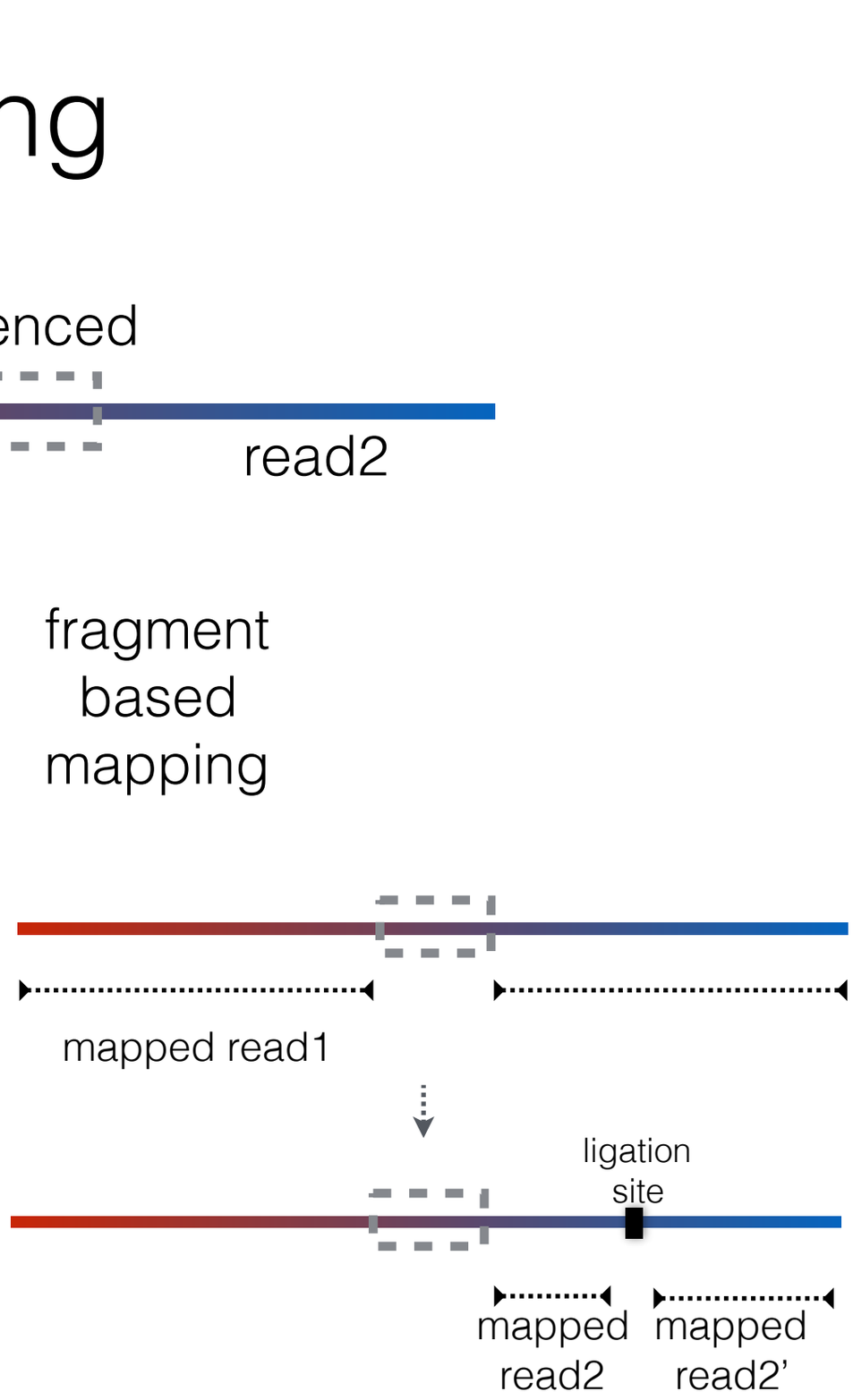
HiC mapping



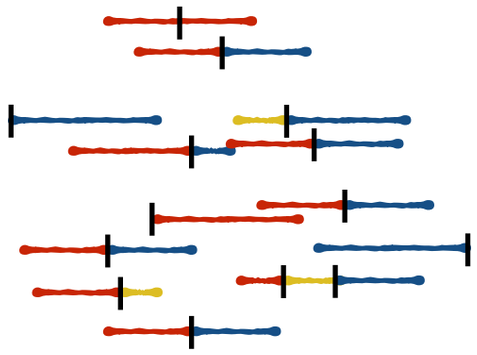


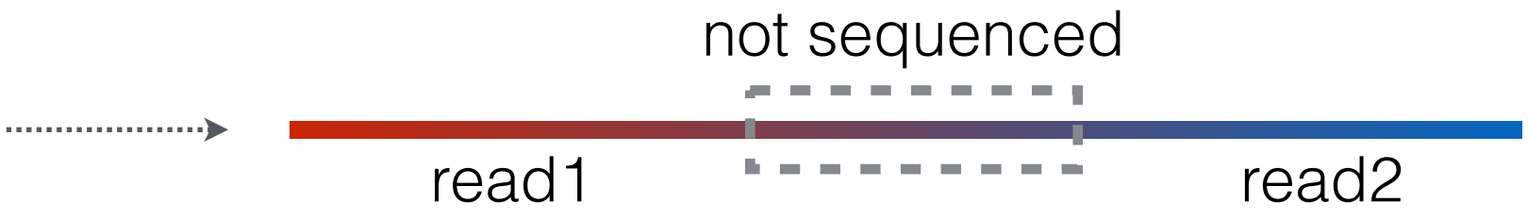
Iterative mapping

fragment based mapping



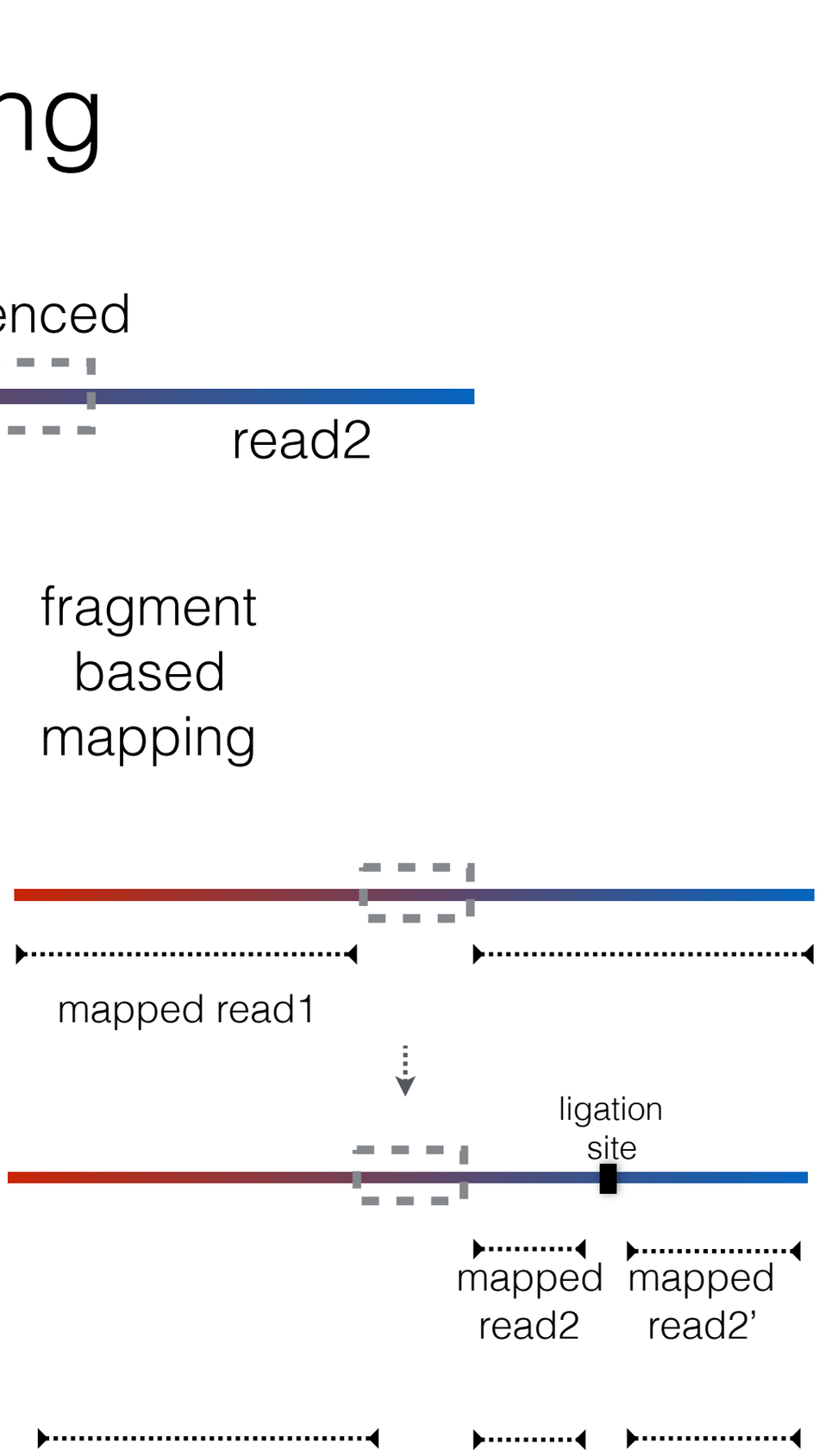
HiC mapping



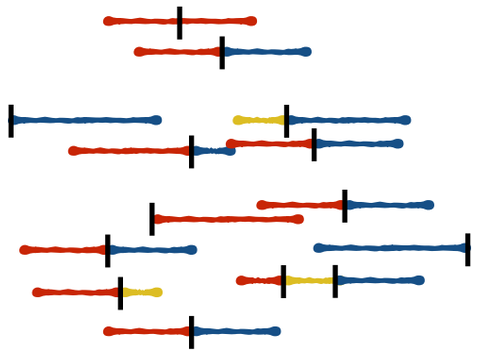


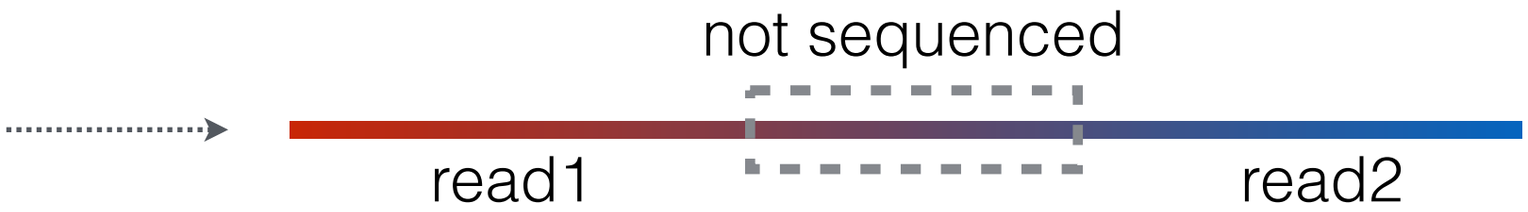
Iterative mapping

fragment based mapping



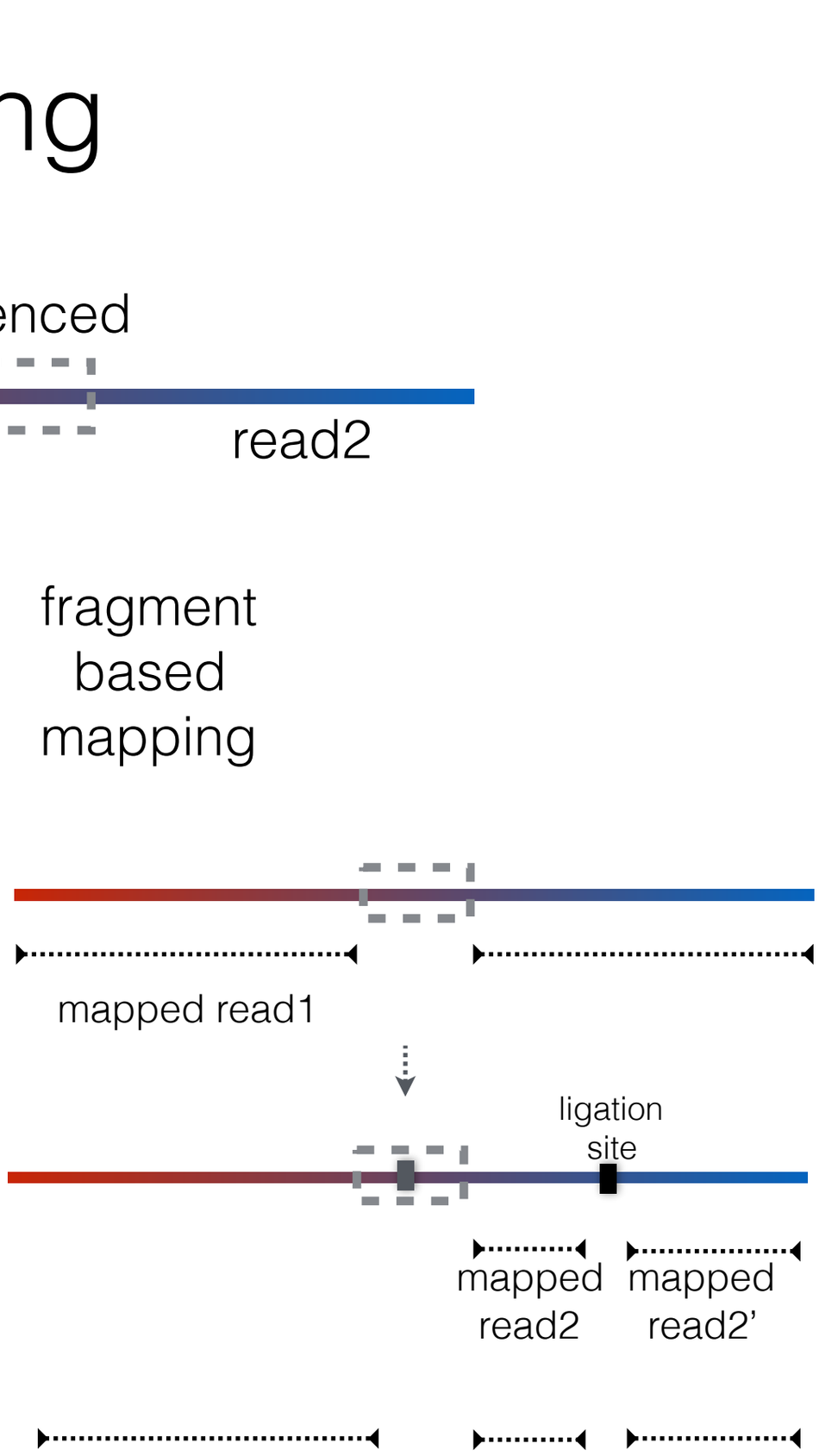
HiC mapping



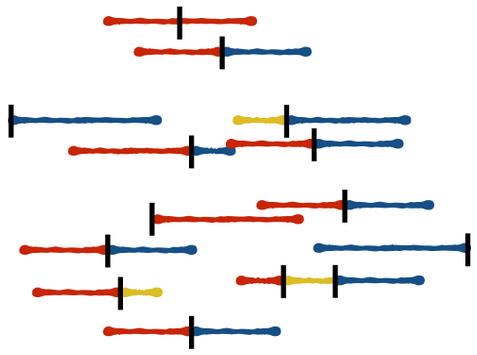


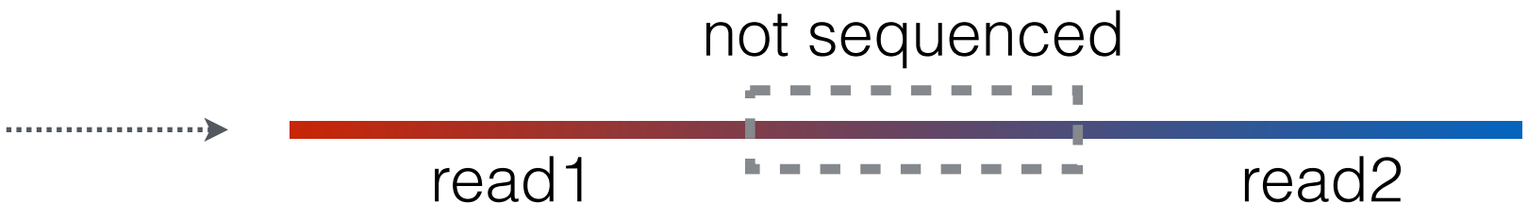
Iterative mapping

fragment based mapping



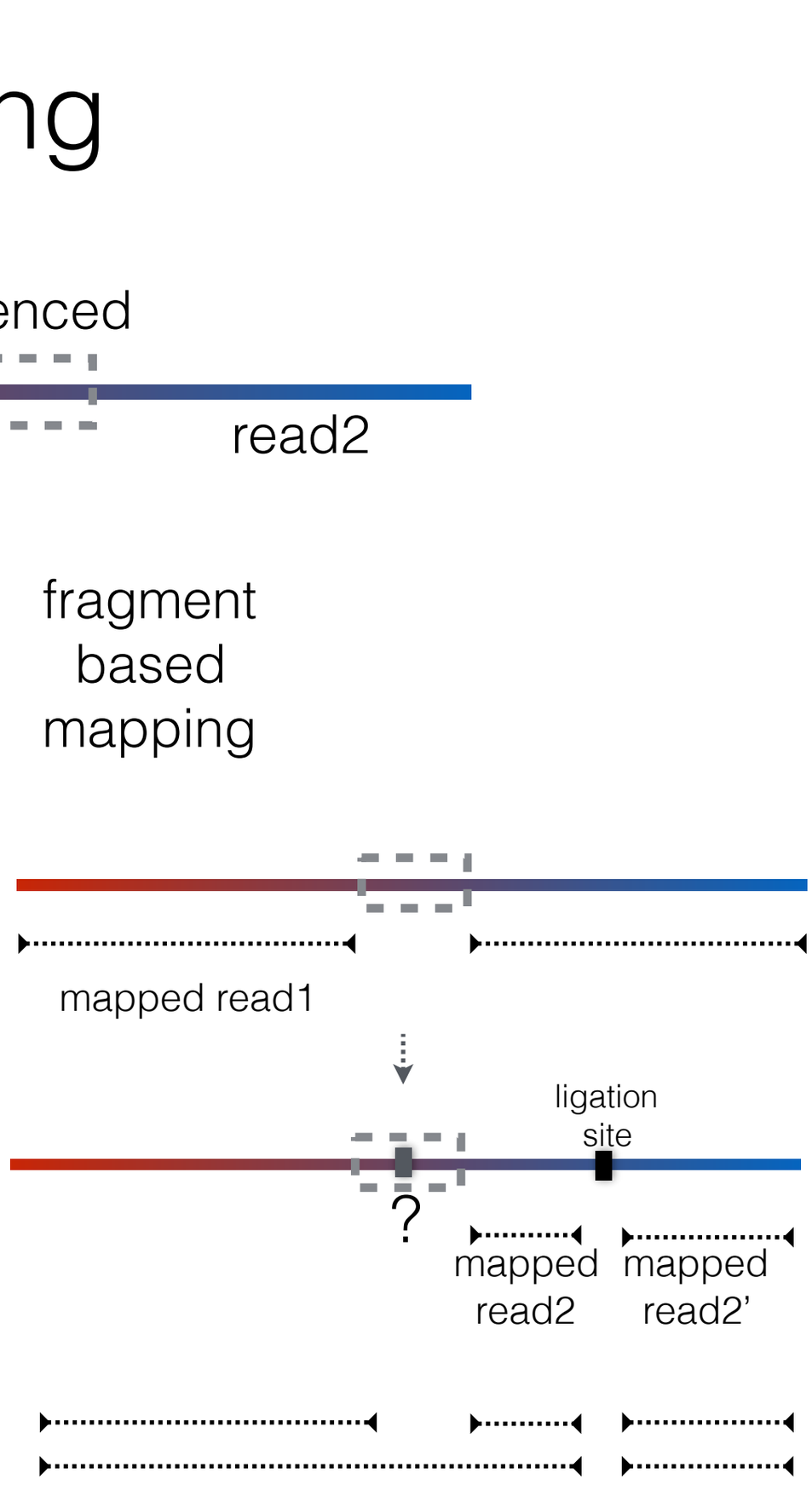
HiC mapping





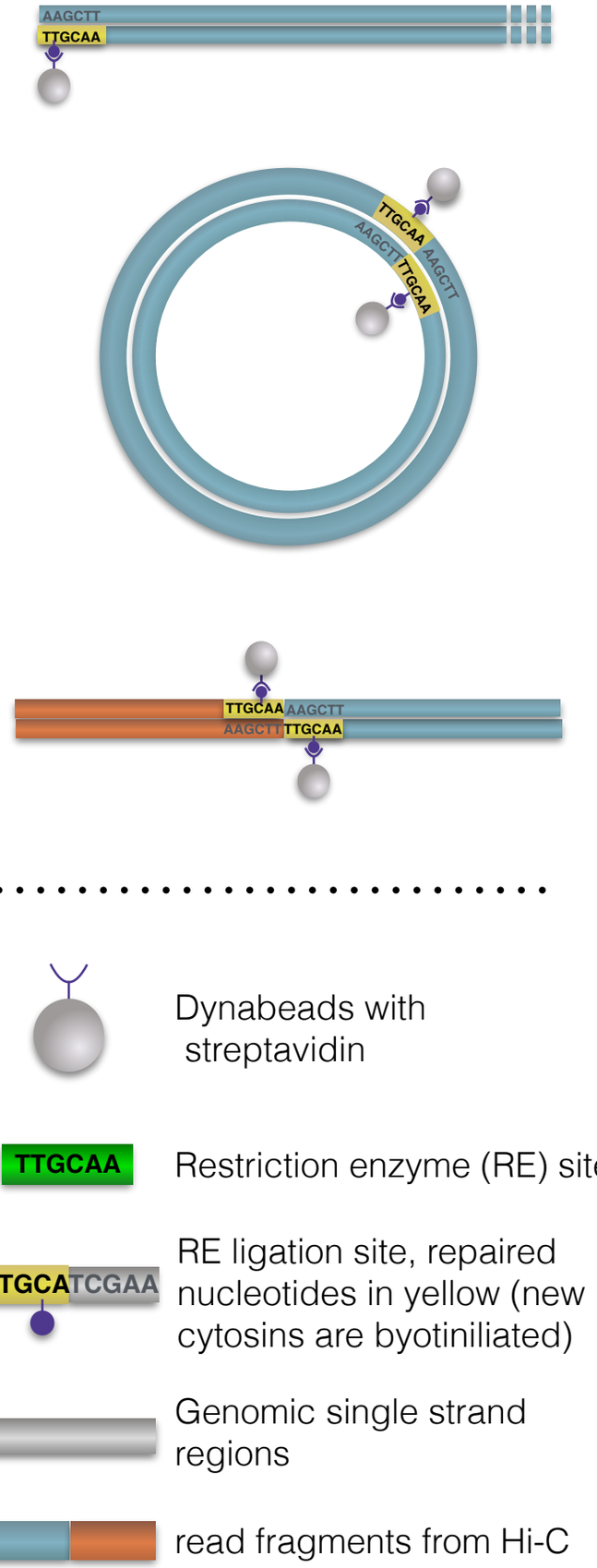
Iterative mapping

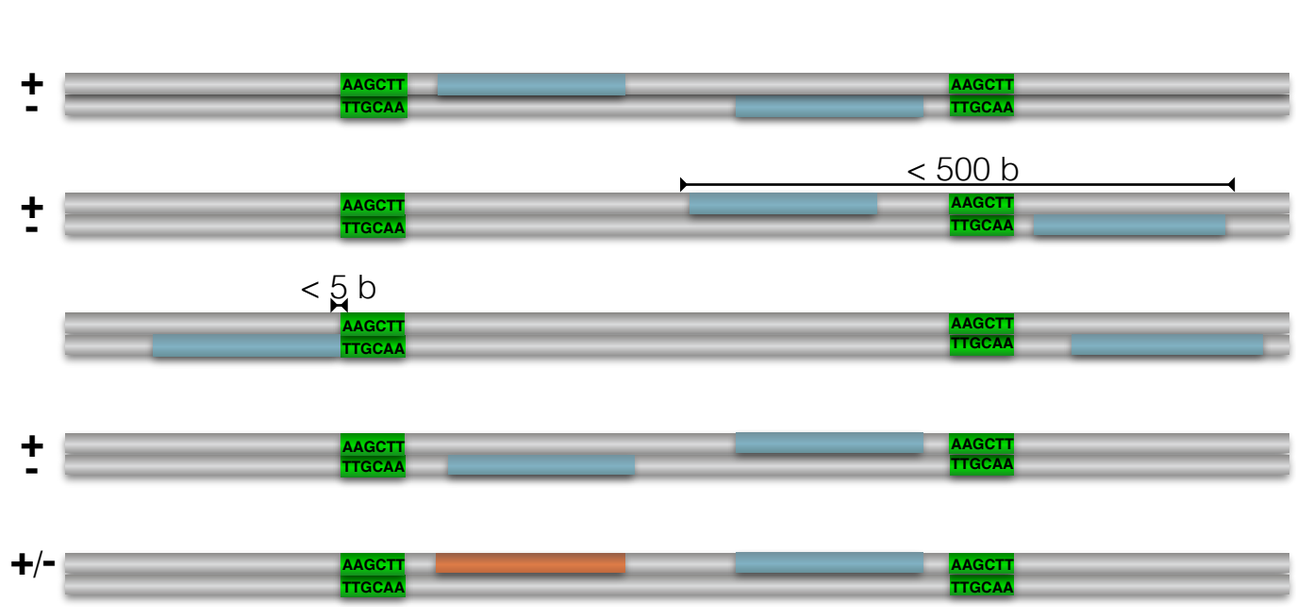
fragment based mapping



How much do we map?

- 80-90% each end => 64-81% intersection
- 1% multiple contacts
- many of these will be lost in the filtering…


AAGCTT
TTGCAA
TTGCAA
AAGCTT
AAGCTT
TTGCAA
TTGCAA
AAGCTT
AAGCTT
TTGCAA
Dynabeads with streptavidin
TTGCAA
Restriction enzyme (RE) site
TTGCATCGAA
RE ligation site, repaired nucleotides in yellow (new cytosins are byotiniliated)
Genomic single strand regions
read fragments from Hi-C



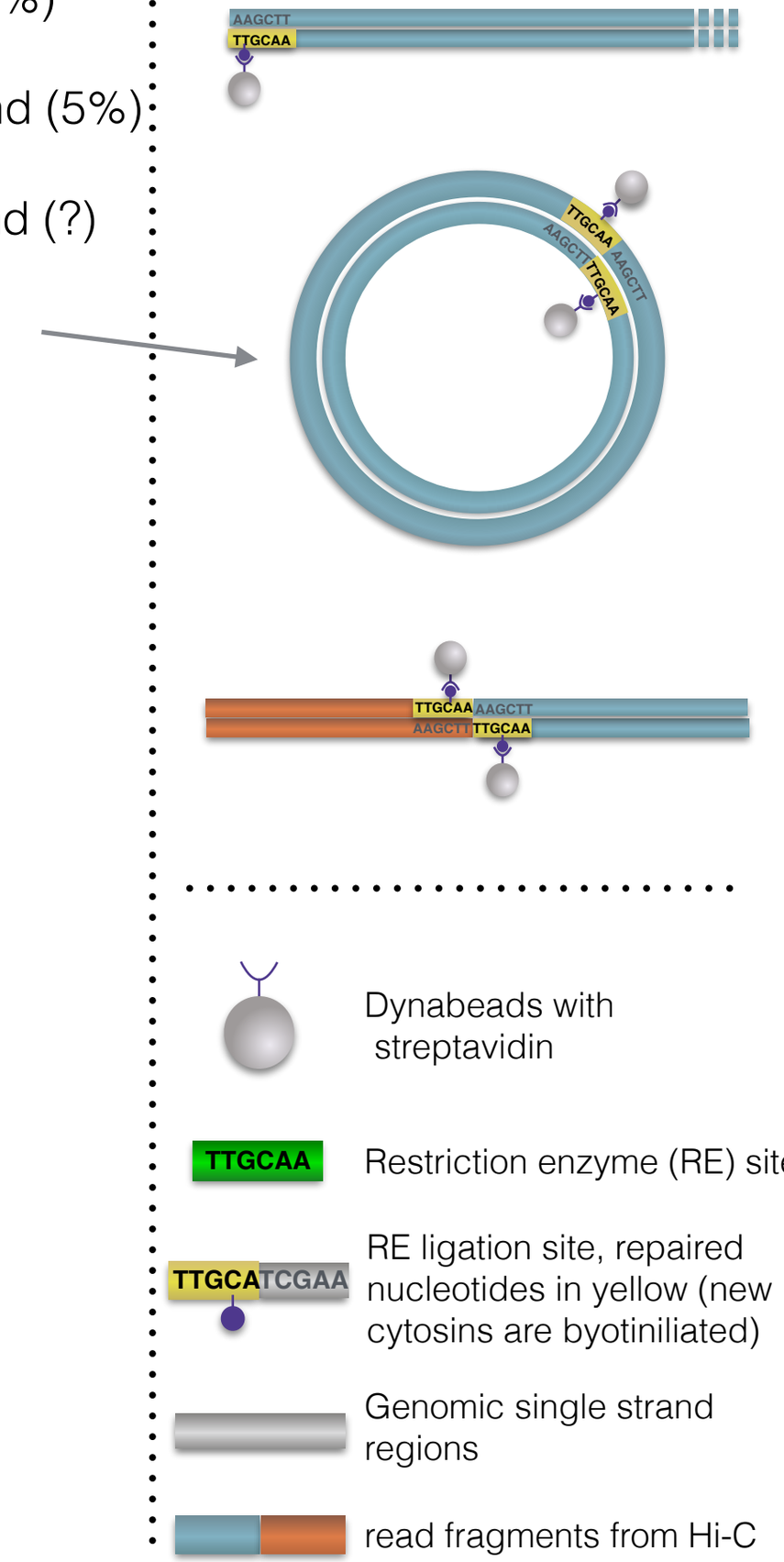
Dangling-end (15%)

Extra dangling-end (5%)

Semi dangling-end (?)

Self-circle (10%)

Error (<2%)



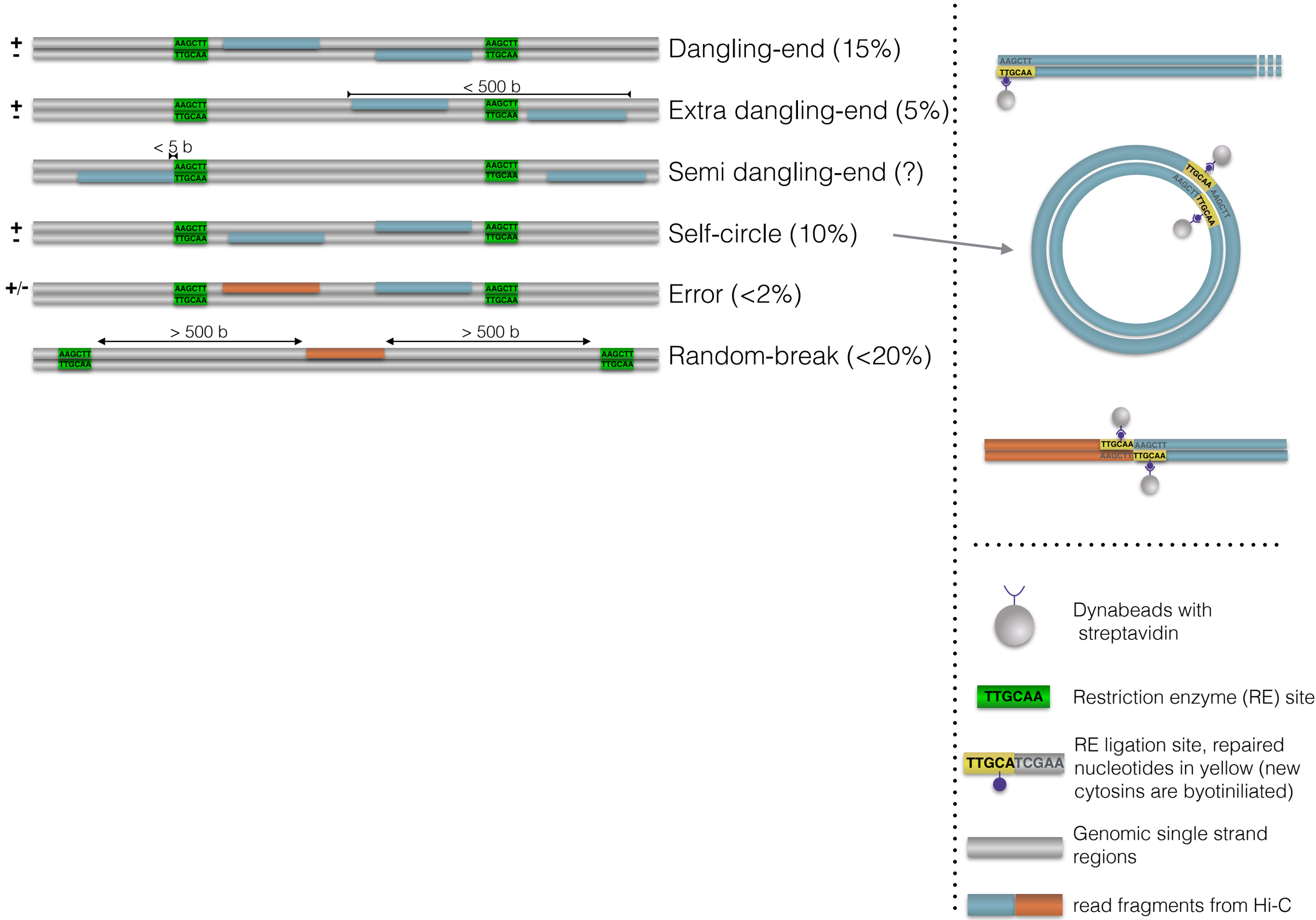
Dynabeads with streptavidin

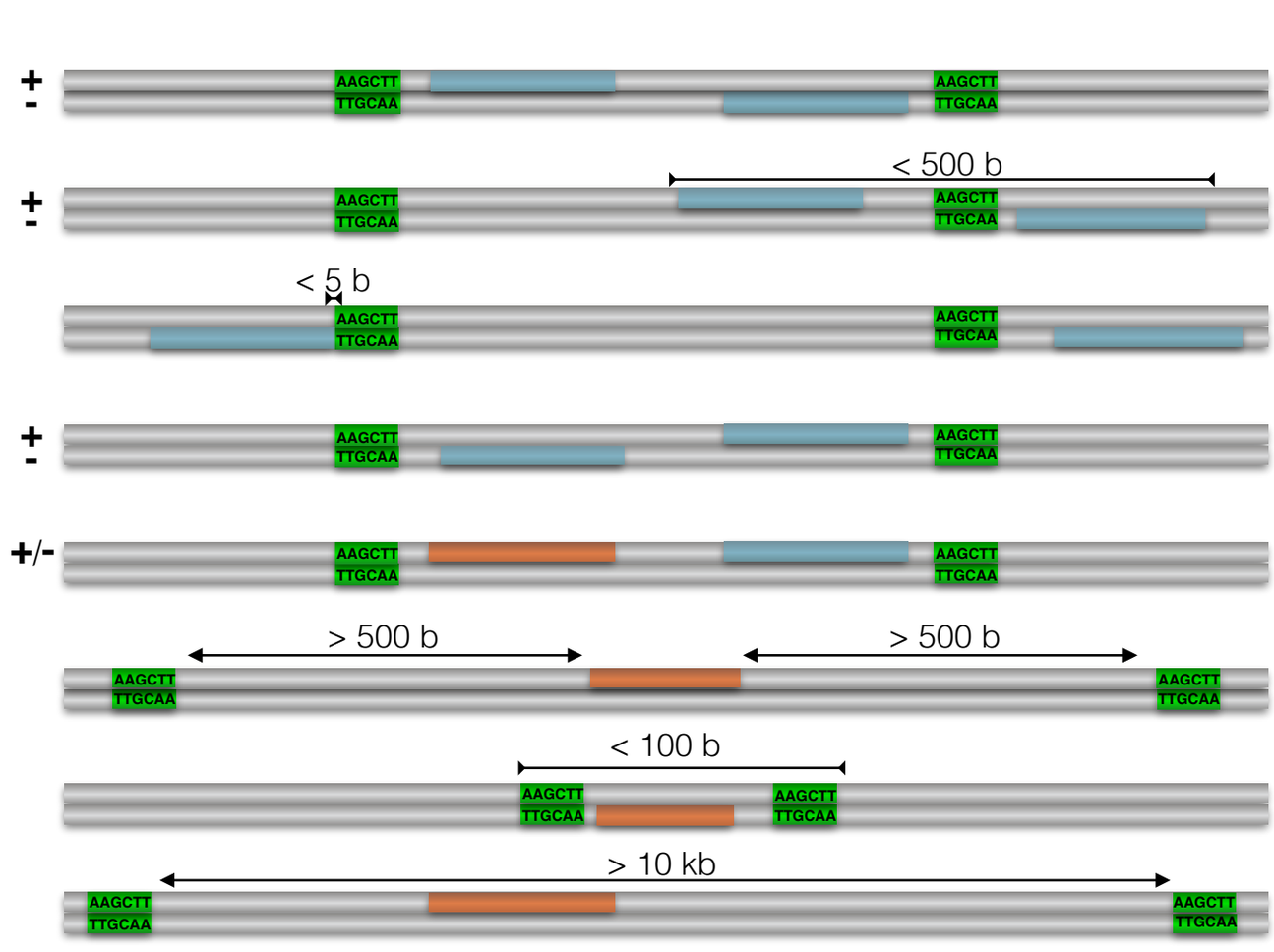
TTGCAA Restriction enzyme (RE) site

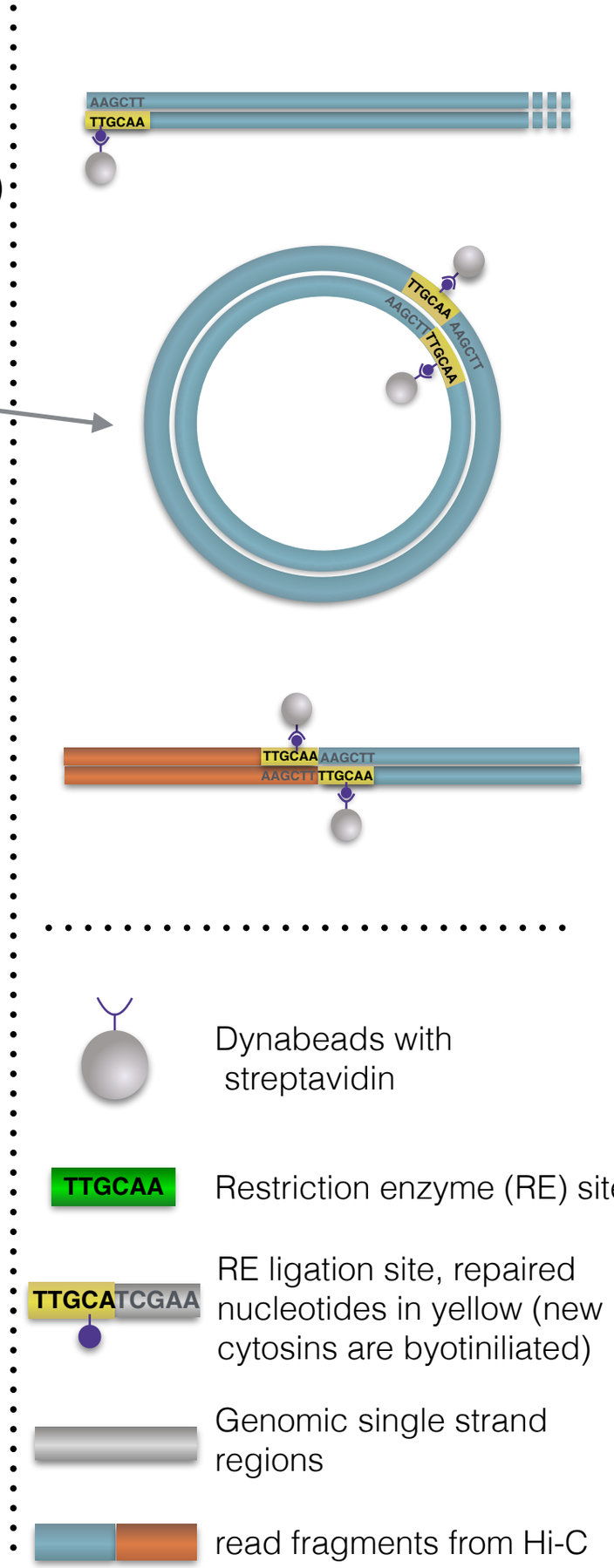
TTGCATCGAA RE ligation site, repaired nucleotides in yellow (new cytosins are byotiniliated)

Genomic single strand regions

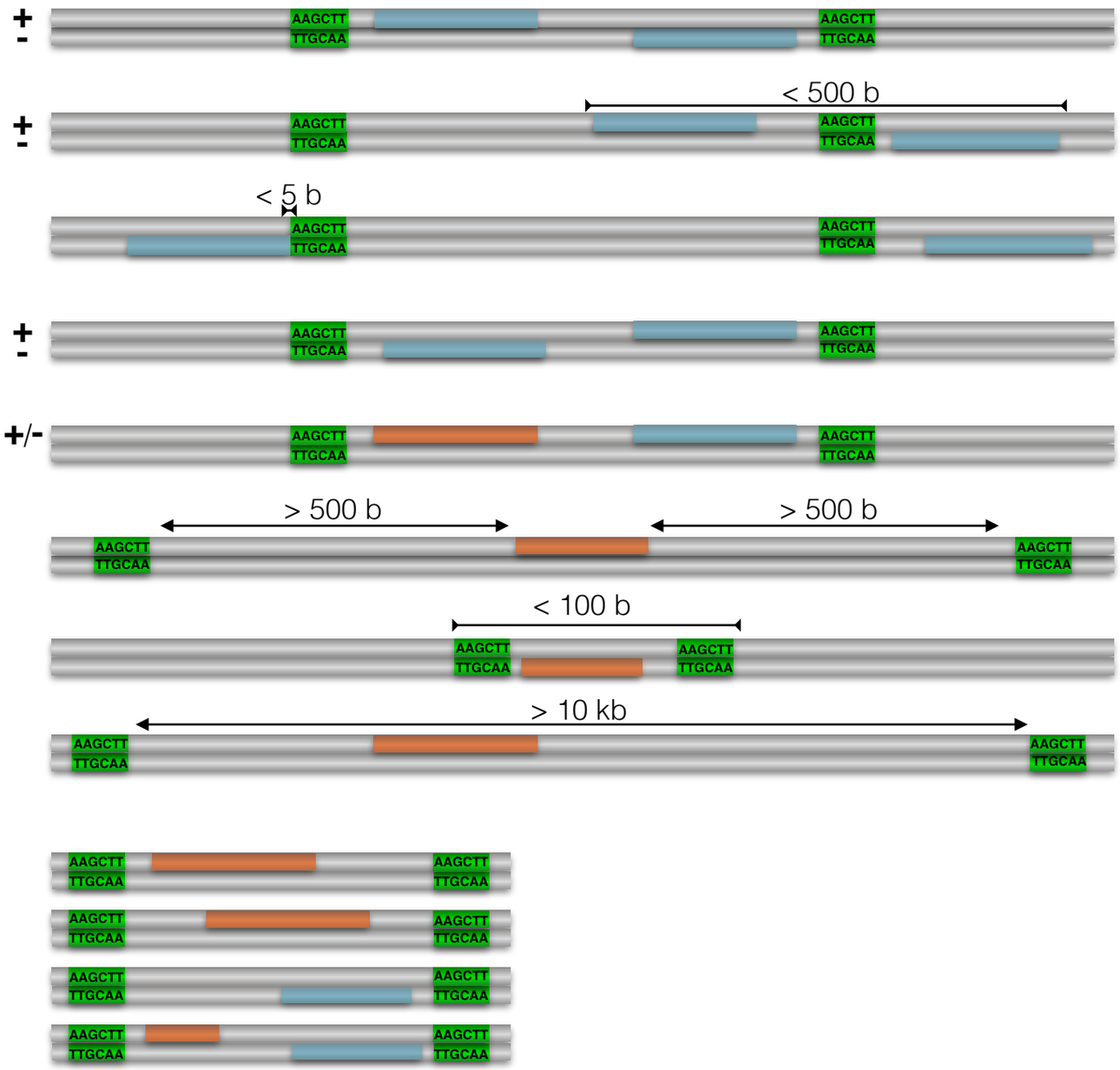
read fragments from Hi-C

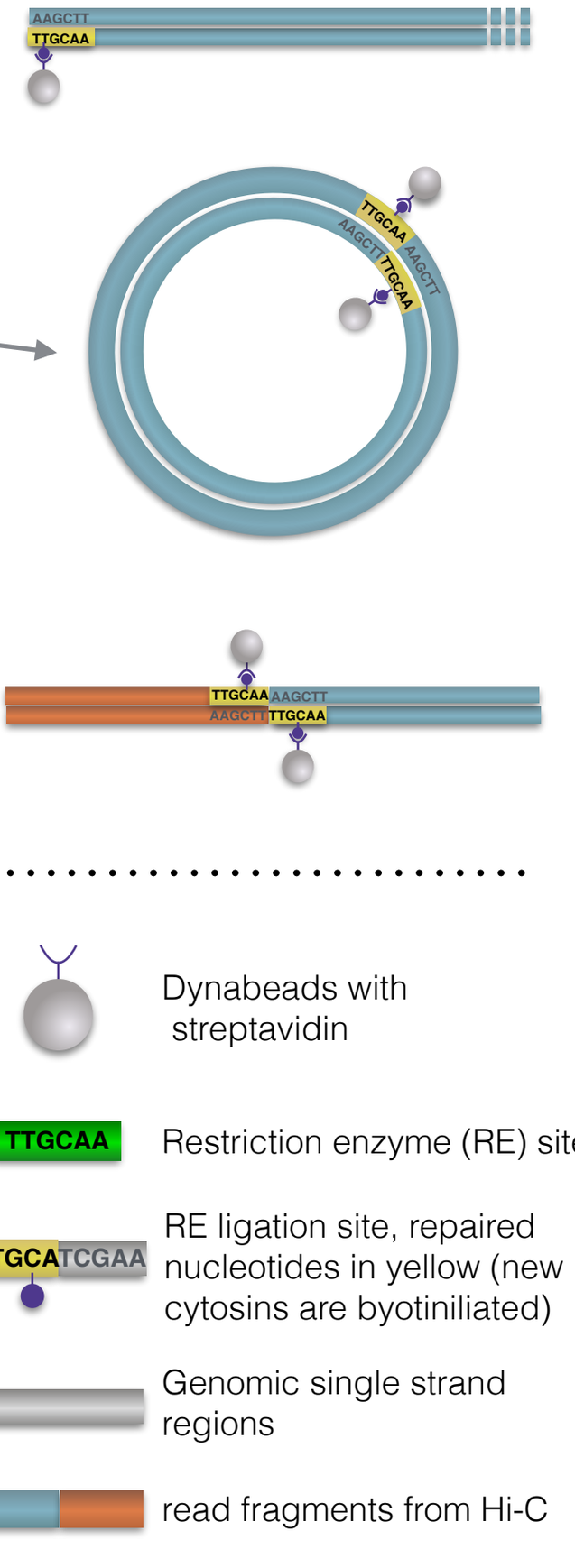

+
-
AAGCTT
TTGCAA
Dangling-end (15%)
< 500 b
Extra dangling-end (5%)
< 5 b
Semi dangling-end (?)
Self-circle (10%)
+/-
Error (<2%)
> 500 b
> 500 b
Random-break (<20%)
TTGCAA
AAGCTT
Dynabeads with streptavidin
Restriction enzyme (RE) site
TTGCATCGAA
RE ligation site, repaired nucleotides in yellow (new cytosins are byotiniliated)
Genomic single strand regions
read fragments from Hi-C


+
-
AAGCTT
TTGCAA
Dangling-end (15%)
< 500 b
Extra dangling-end (5%)
< 5 b
Semi dangling-end (?)
Self-circle (10%)
+/-
Error (<2%)
> 500 b
Random-break (<20%)
< 100 b
Too short (<1%)
> 10 kb
Too large (<1%)


AAGCTT
TTGCAA
TTGCAA
AAGCTT
AAGCTT
TTGCAA
TTGCAA
AAGCTT
Dynabeads with streptavidin
TTGCAA
Restriction enzyme (RE) site
TTGCATCGAA
RE ligation site, repaired nucleotides in yellow (new cytosins are byotiniliated)
Genomic single strand regions
read fragments from Hi-C

+ -
AAGCTT
TTGCAA
Dangling-end (15%)
+ -
< 500 b
Extra dangling-end (5%)
< 5 b
Semi dangling-end (?)
+ -
Self-circle (10%)
+/-
Error (<2%)
> 500 b
> 500 b
Random-break (<20%)
< 100 b
Too short (<1%)
> 10 kb
Too large (<1%)
Over-represented (<1%)
Duplicated (20%)
Dynabeads with streptavidin
TTGCAA
Restriction enzyme (RE) site
TTGCATCGAA
RE ligation site, repaired nucleotides in yellow (new cytosins are byotiniliated)
Genomic single strand regions
read fragments from Hi-C


+
-
AAGCTT
TTGCAA
Dangling-end (15%)
< 500 b
Extra dangling-end (5%)
< 5 b
Semi dangling-end (?)
Self-circle (10%)
+/-
Error (<2%)
> 500 b
Random-break (<20%)
< 100 b
Too short (<1%)
> 10 kb
Too large (<1%)
Over-represented (<1%)
Duplicated (20%)
Valid pair
(60% of intersection,
<50% of the total)


AAGCTT
TTGCAA
TTGCAA
AAGCTT
AAGCTT
TTGCAA
TTGCAA
AAGCTT
TTGCAA
AAGCTT
Dynabeads with
streptavidin
TTGCAA
Restriction enzyme (RE) site
TTGCATCGAA
RE ligation site, repaired
nucleotides in yellow (new
cytosins are byotiniliated)
Genomic single strand
regions
read fragments from Hi-C