CRAC, a software for analyzing RNA-seq reads

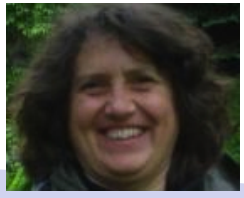
N. Philippe

T. Commes

É. Rivals

M. Salson

CRAC, a software for analyzing RNA-seq reads



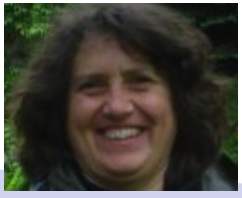
N. Philippe **T. Commes**

Institut de recherche en biothérapie
Inserm – Univ. Montpellier 1 – CHU

É. Rivals

M. Salson

CRAC, a software for analyzing RNA-seq reads



N. Philippe

T. Commes

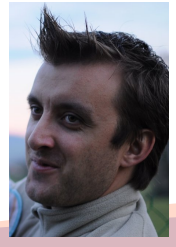
É. Rivals

Institut de recherche en biothérapie
Inserm – Univ. Montpellier 1 – CHU

LIRMM – Équipe MAB
CNRS – Univ. Montpellier 2

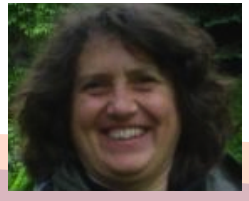
M. Salson

CRAC, a software for analyzing RNA-seq reads



N. Philippe

Institut de biologie computationnelle



T. Commes

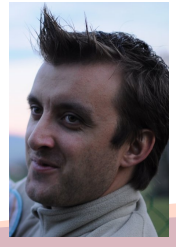
É. Rivals

LIRMM – Équipe MAB
CNRS – Univ. Montpellier 2

Institut de recherche en biothérapie
Inserm – Univ. Montpellier 1 – CHU

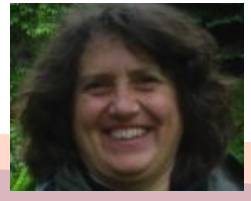
M. Salson

CRAC, a software for analyzing RNA-seq reads



N. Philippe

Institut de biologie computationnelle
Institut de recherche en biothérapie
Inserm – Univ. Montpellier 1 – CHU



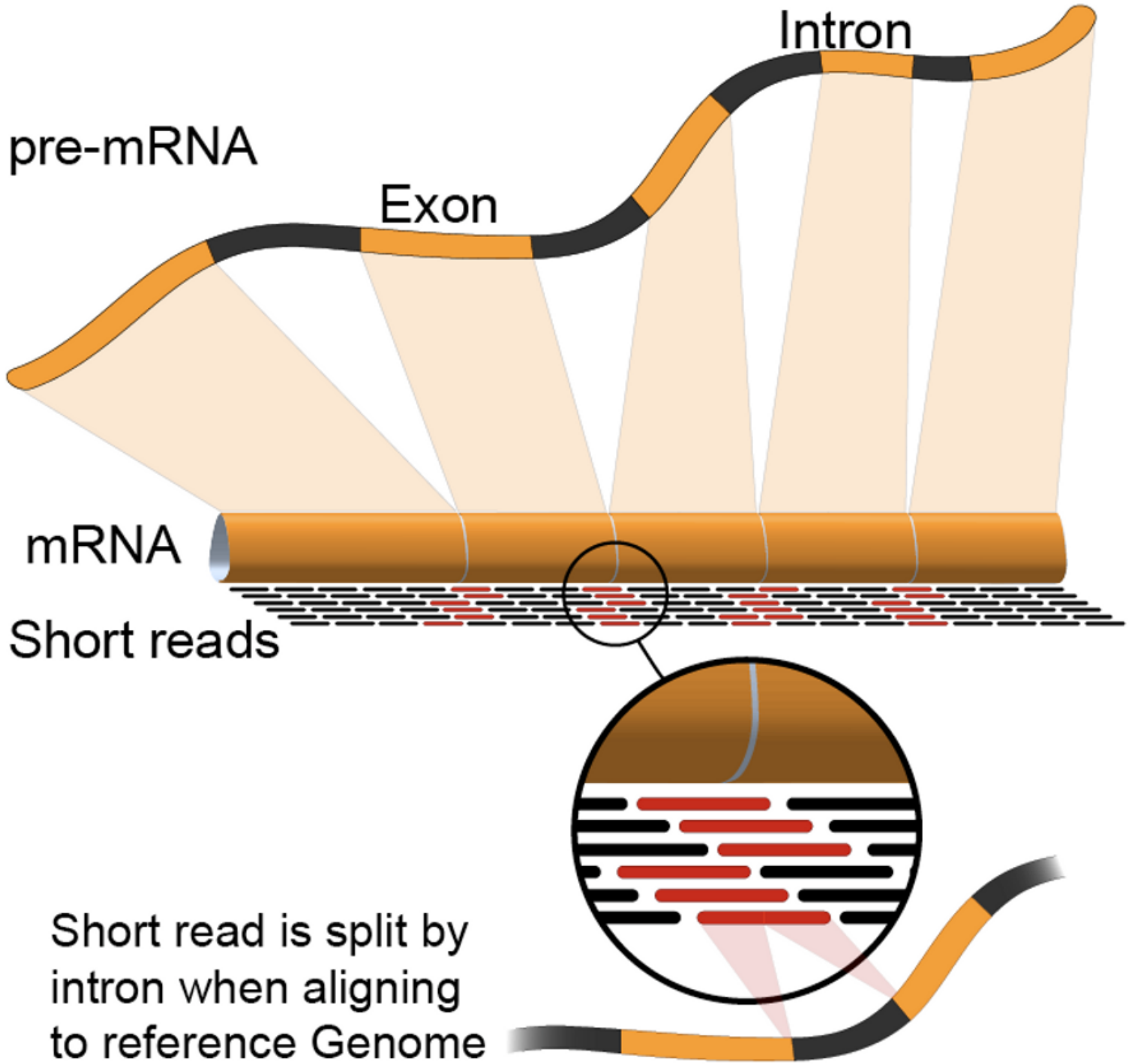
T. Commes

É. Rivals

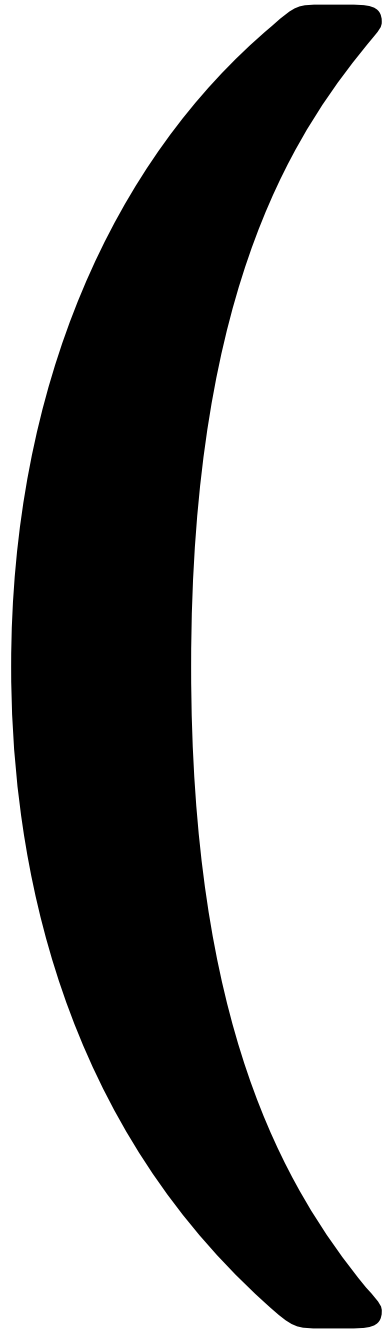
LIRMM – Équipe MAB
CNRS – Univ. Montpellier 2

M. Salson

Équipe Bonsai
LIFL (CNRS – Univ. Lille 1)
Inria


Intron
pre-mRNA
Exon
mRNA
Short reads
Short read is split by
intron when aligning
to reference Genome

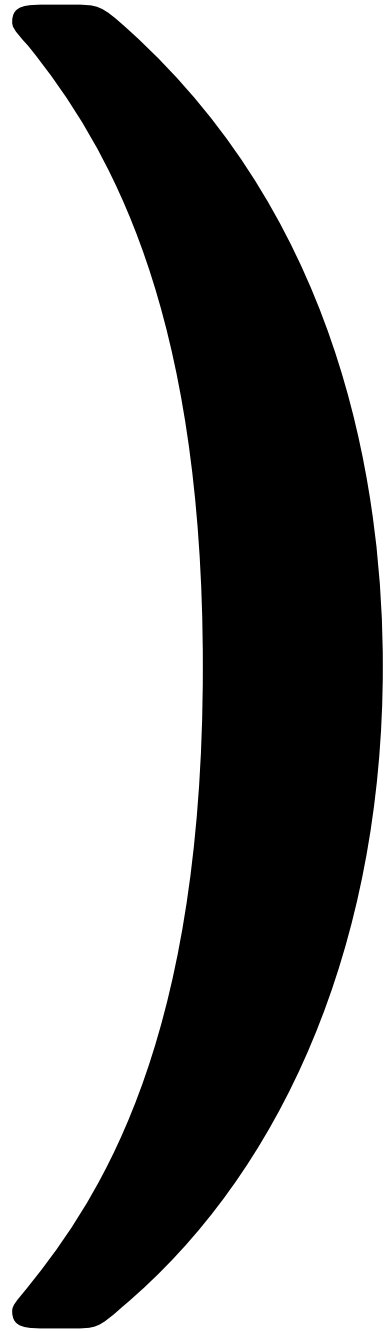

CGACTAGCTAGCATCAGCATCTATTATAGCATCATCGACTATACGACTATCGATCATCGTATATATAGGAGATCACAG



Underlying hypothesis

Underlying hypothesis

1 read $\longrightarrow$ 1 location



In our bibliography

In our bibliography

Using reads to annotate the genome: influence of length, background distribution, and sequence errors on prediction capacity

Nicolas Philippe[1,2], Anthony Boureux[2], Laurent Bréhélin[1], Jorma Tarhio[3], Thérèse Commes[2] and Éric Rivals[1,*]

22

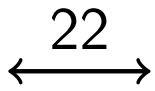
22

for the human genome

22: not too small, not too large

Genome

22: not too small, not too large

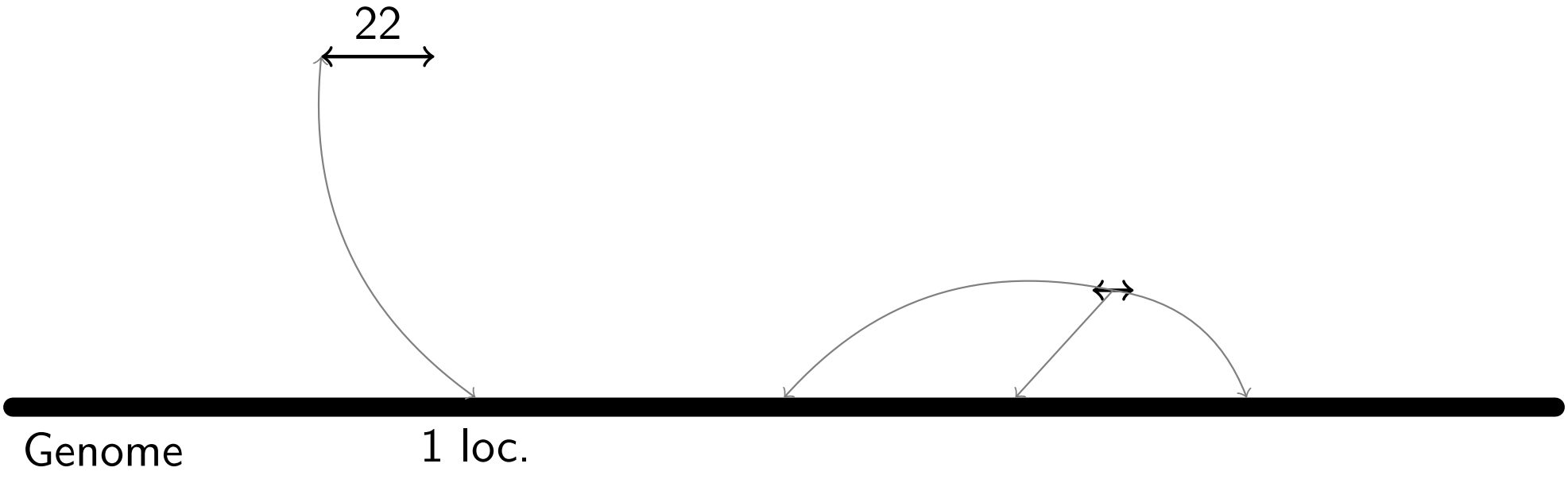


Genome

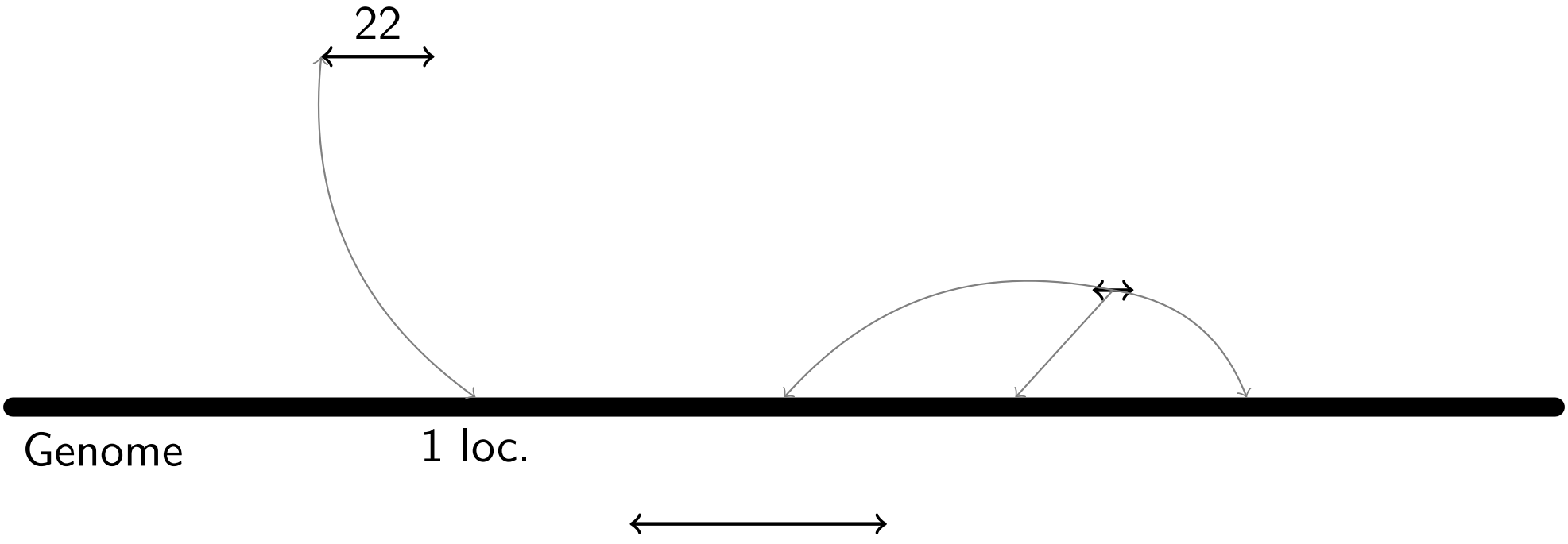
22: not too small, not too large

22: not too small, not too large

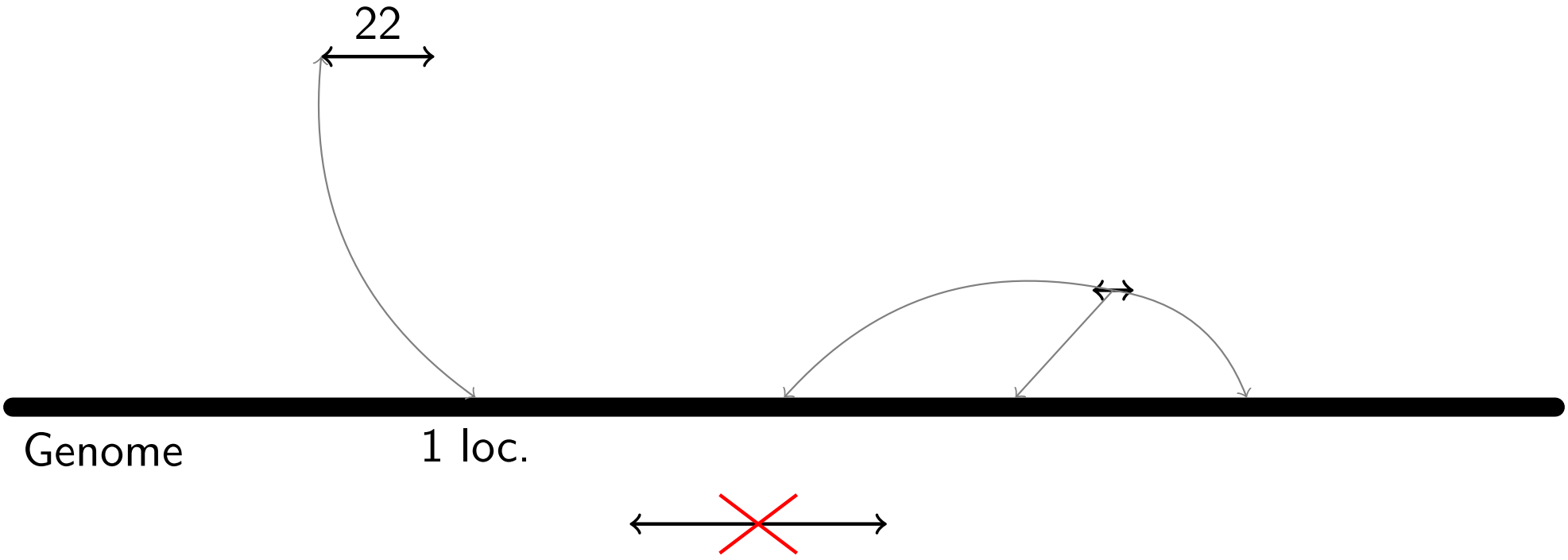
22: not too small, not too large



22: not too small, not too large

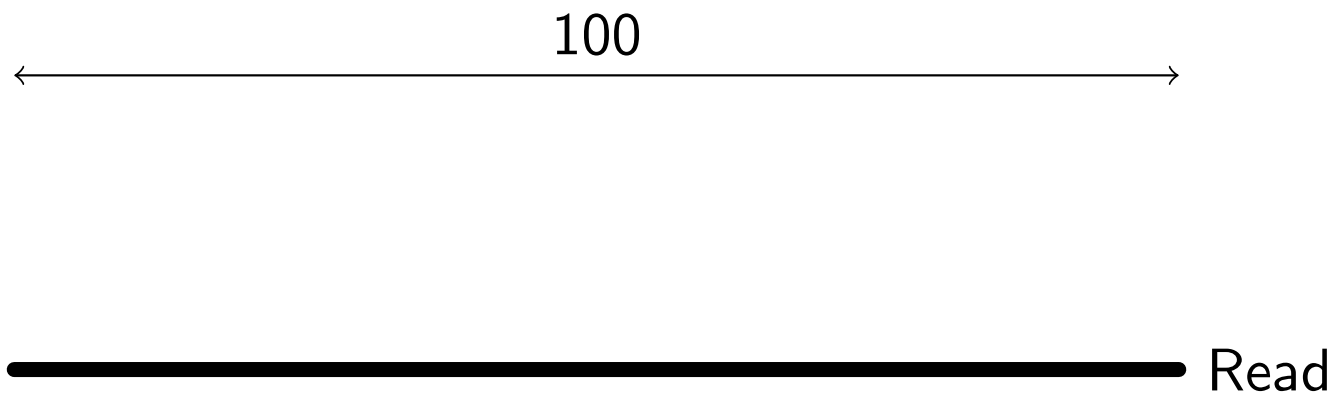


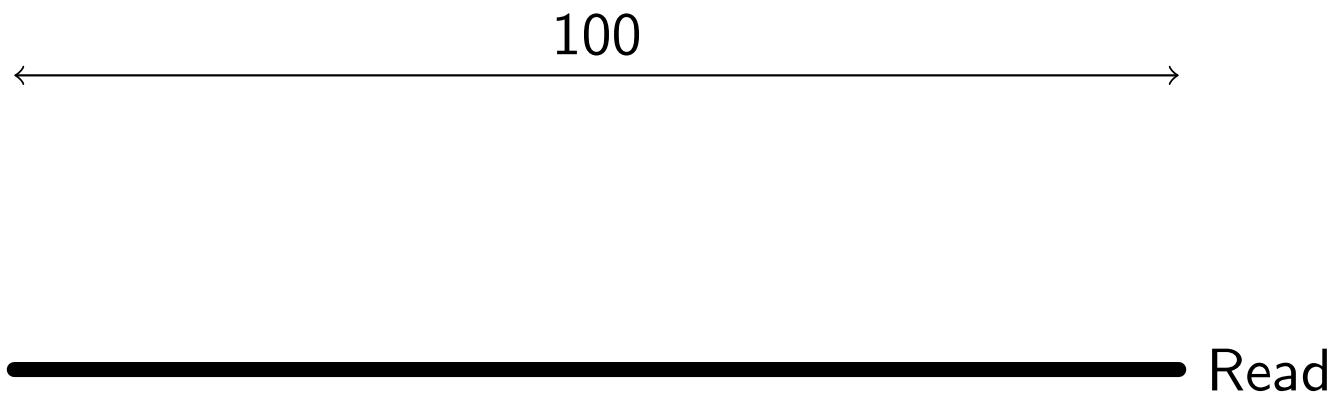
22: not too small, not too large

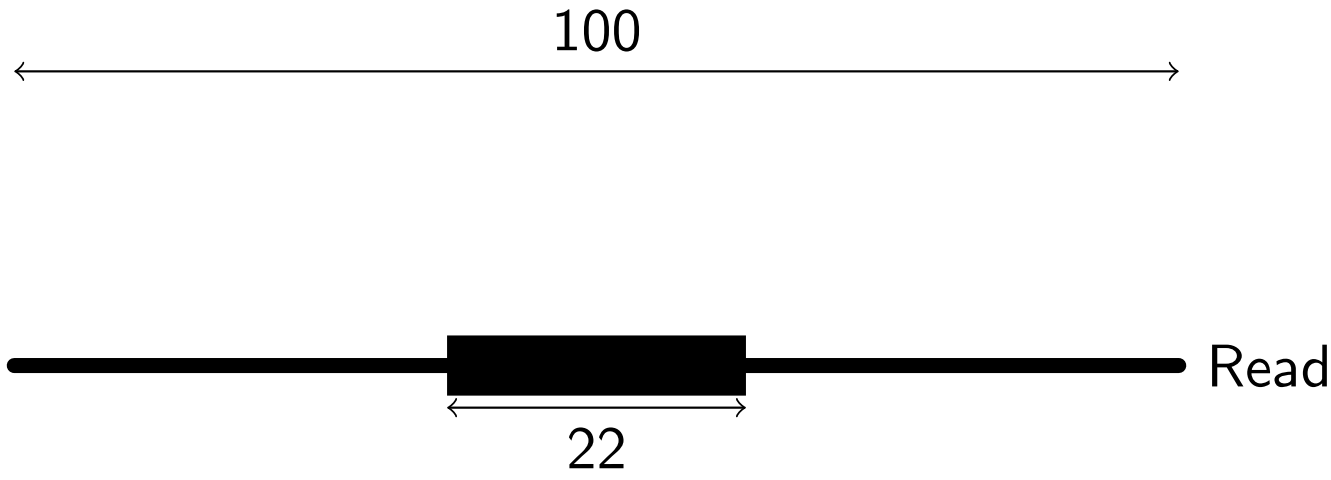


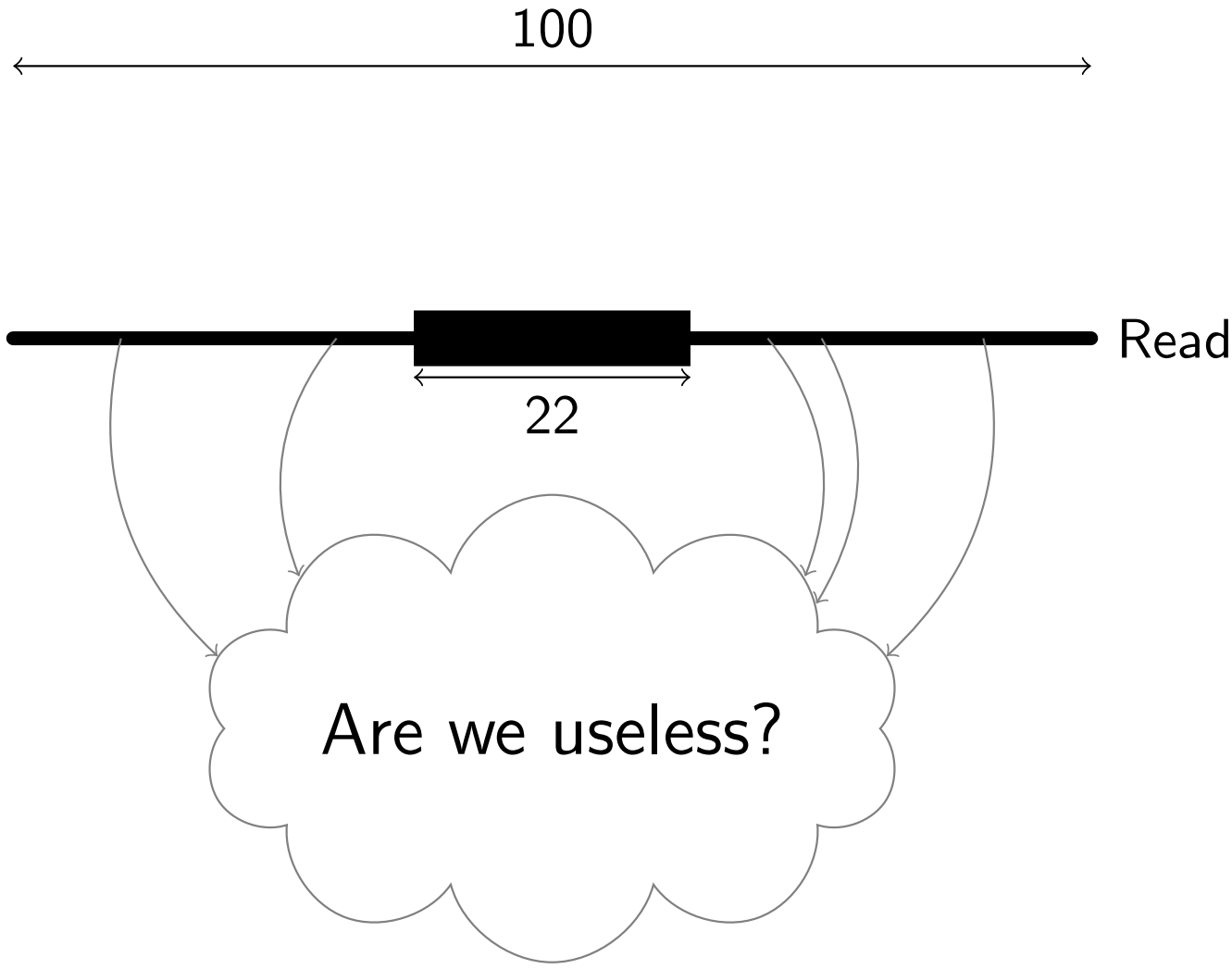
Underlying hypothesis

Resequencing


100
Read


100
Read

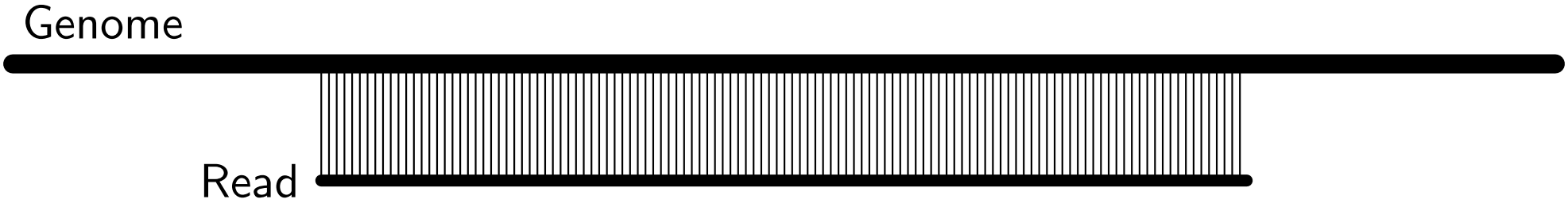

100
Read
22


100
Read
22
Are we useless?

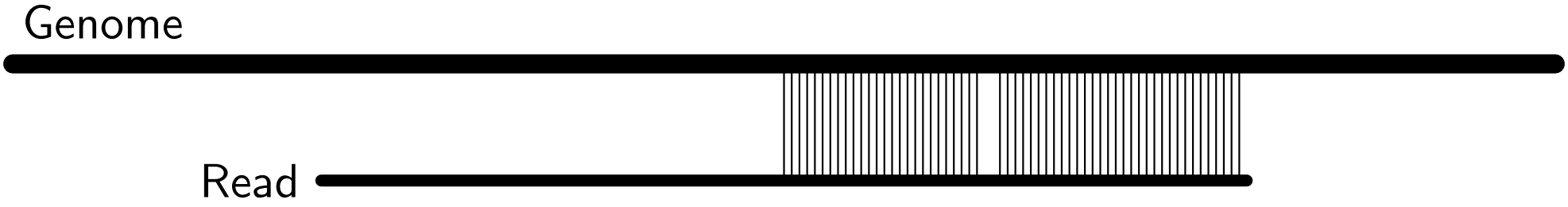
More data

More info

In an ideal world

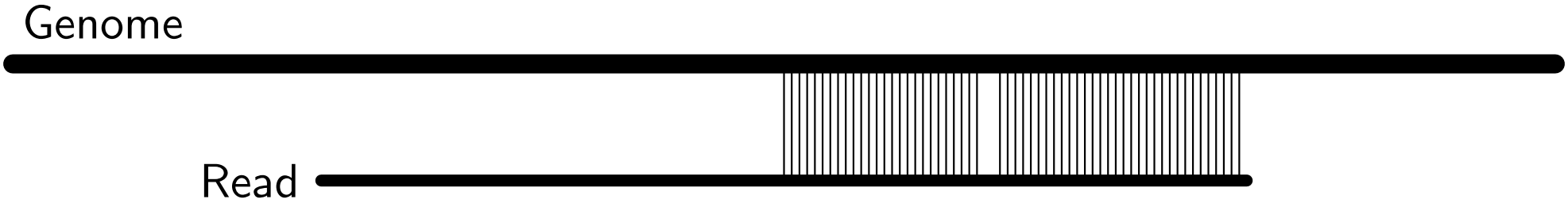


We're not in an ideal world




Why?

We're not in an ideal world

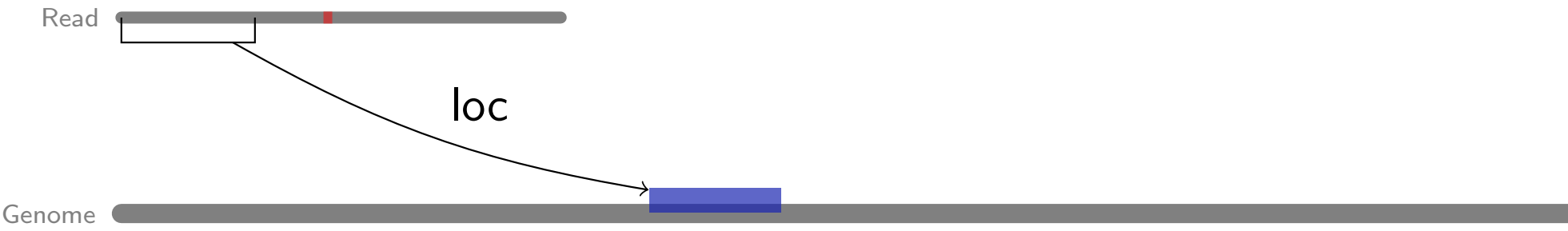


Because!

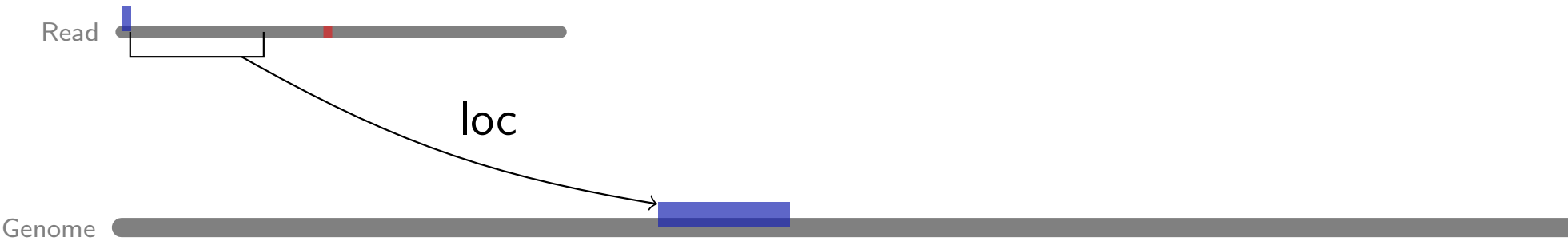
Because!

We can explain what's going on (sometimes)

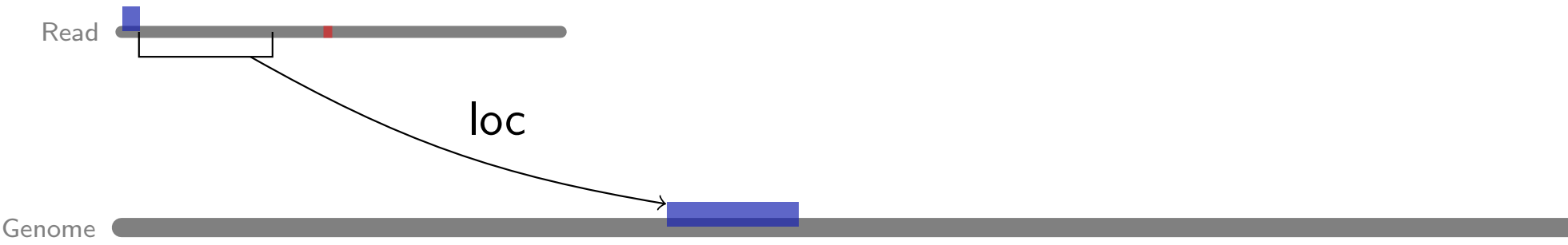
Using 22-mers to decipher a read



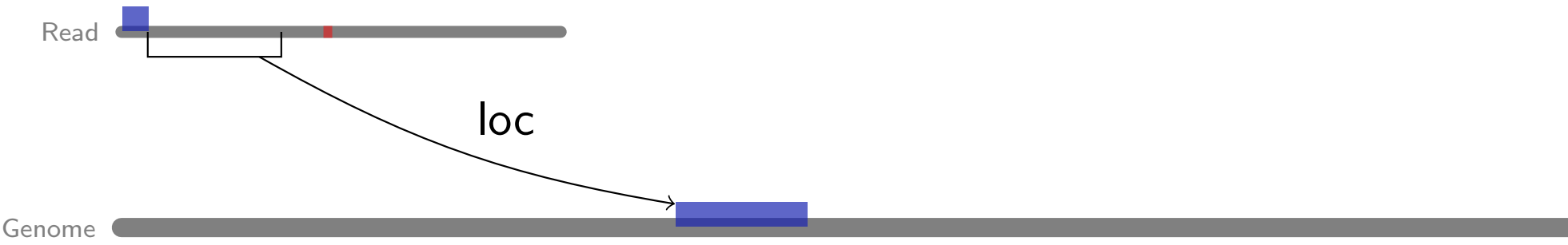
Using 22-mers to decipher a read



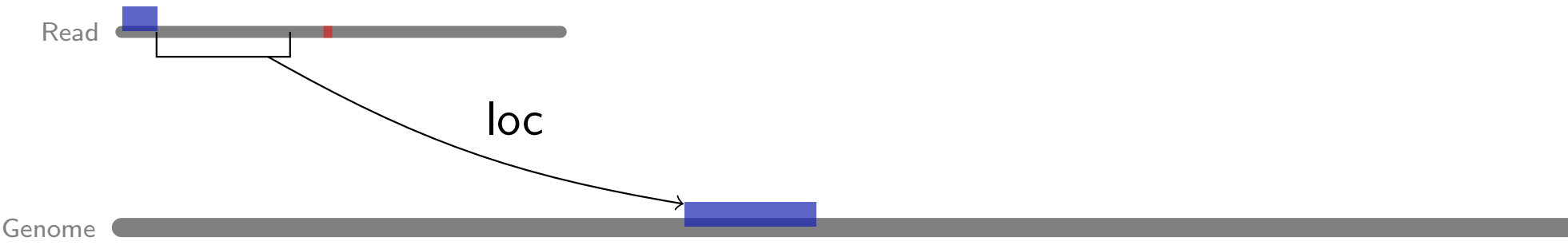
Using 22-mers to decipher a read



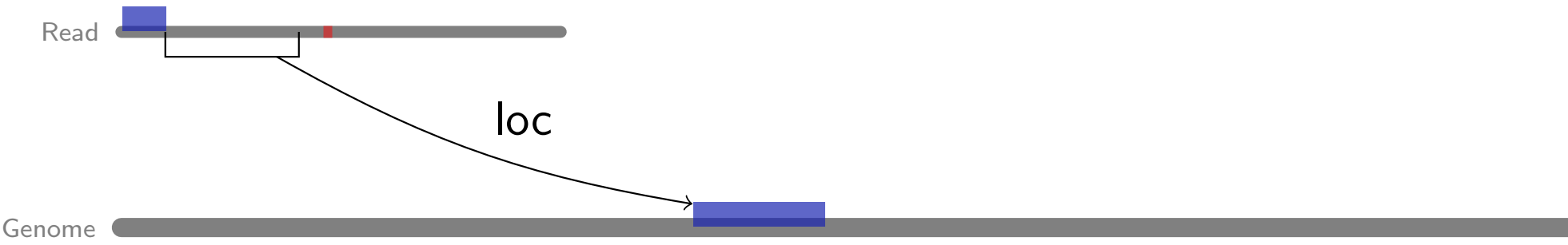
Using 22-mers to decipher a read



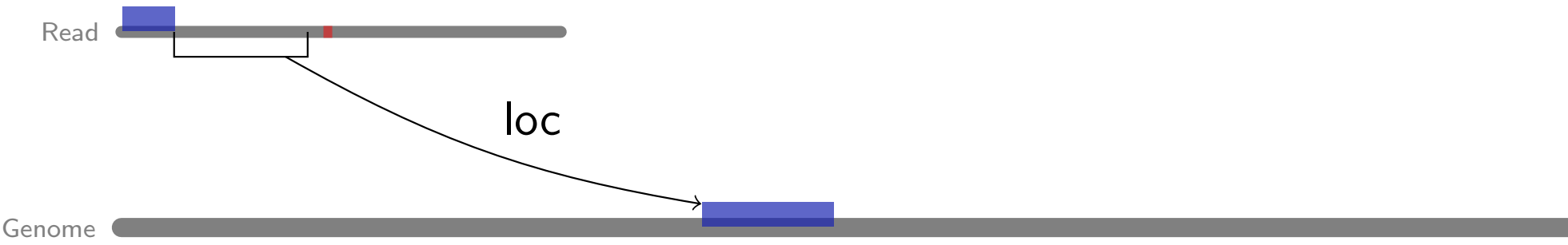
Using 22-mers to decipher a read



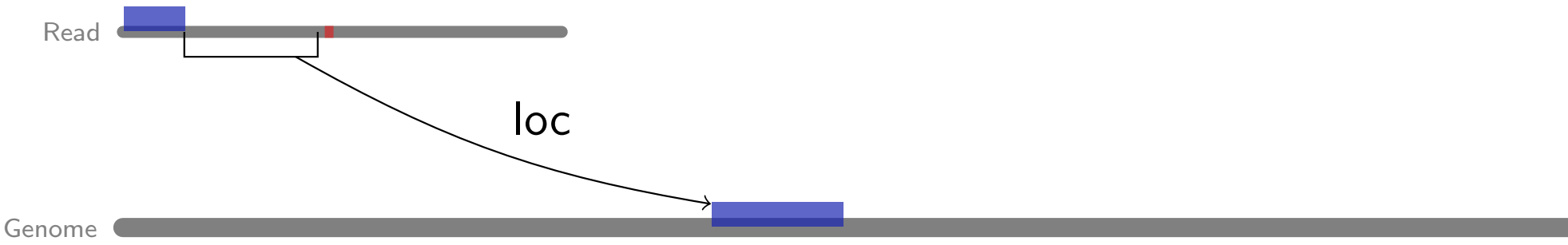
Using 22-mers to decipher a read



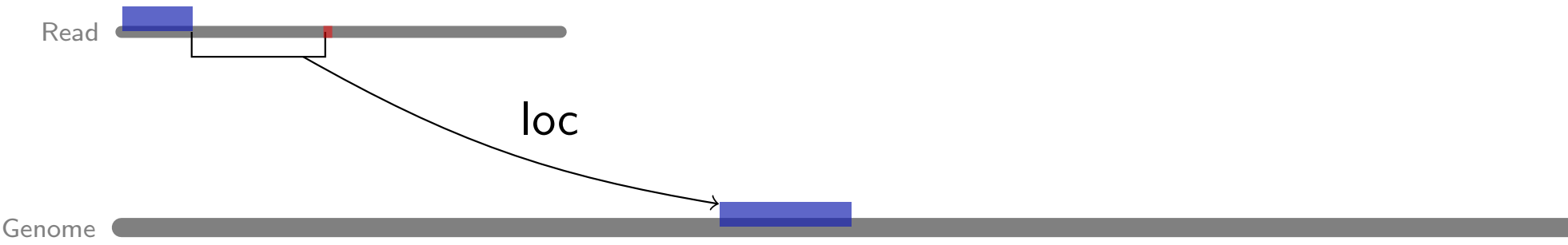
Using 22-mers to decipher a read



Using 22-mers to decipher a read



Using 22-mers to decipher a read



Using 22-mers to decipher a read

Using 22-mers to decipher a read

Using 22-mers to decipher a read

Using 22-mers to decipher a read

Using 22-mers to decipher a read

Using 22-mers to decipher a read

Using 22-mers to decipher a read

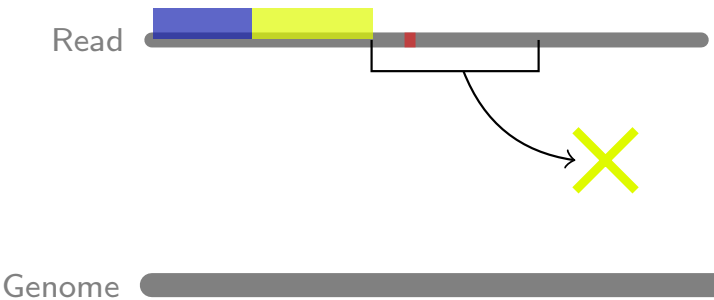
Using 22-mers to decipher a read

Using 22-mers to decipher a read

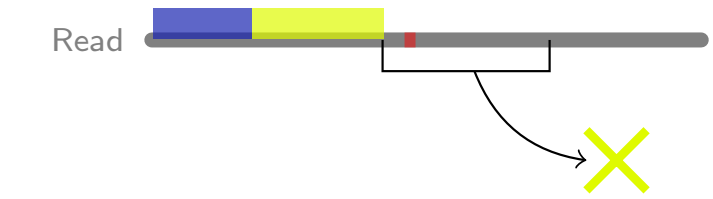
Using 22-mers to decipher a read

Using 22-mers to decipher a read

Using 22-mers to decipher a read

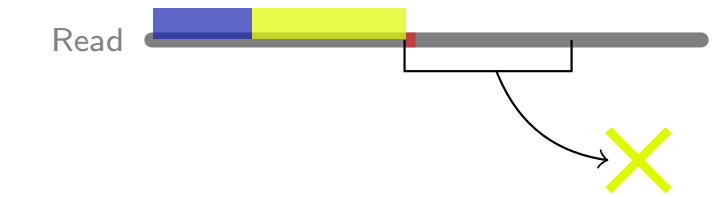


Using 22-mers to decipher a read

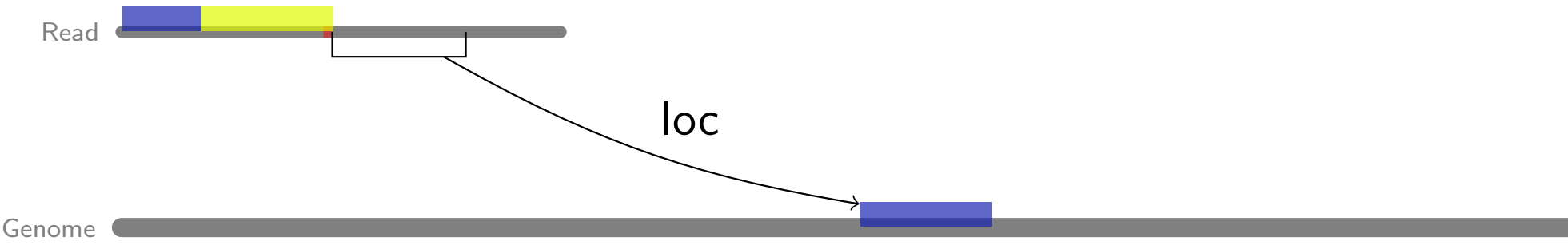


Using 22-mers to decipher a read

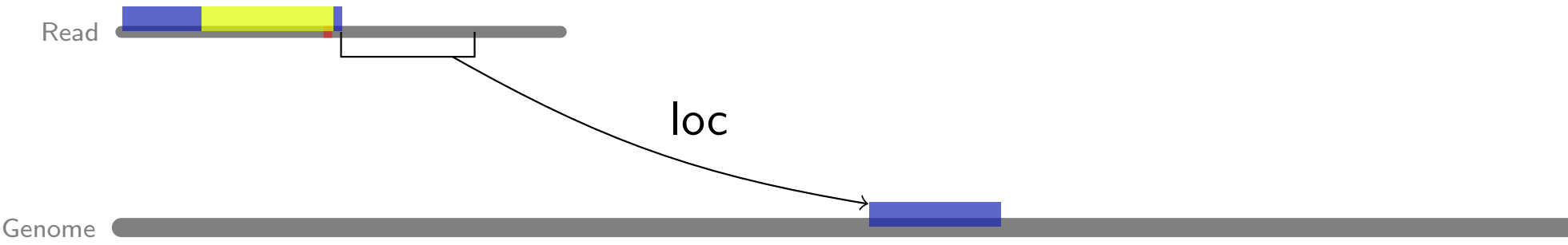
Using 22-mers to decipher a read



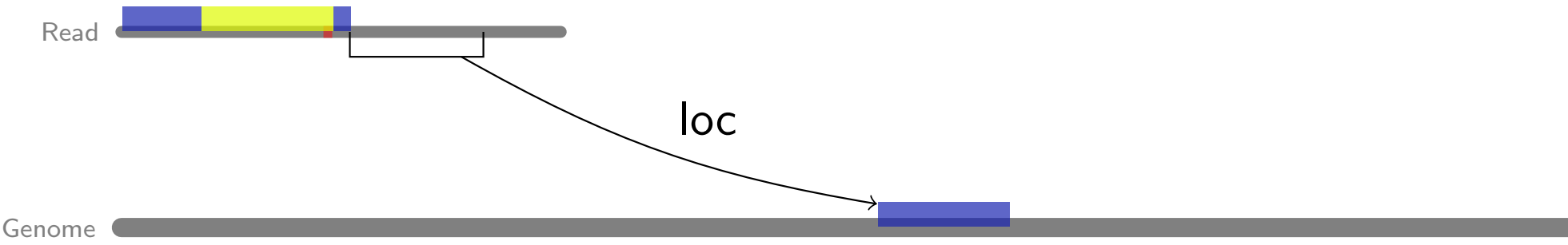
Using 22-mers to decipher a read



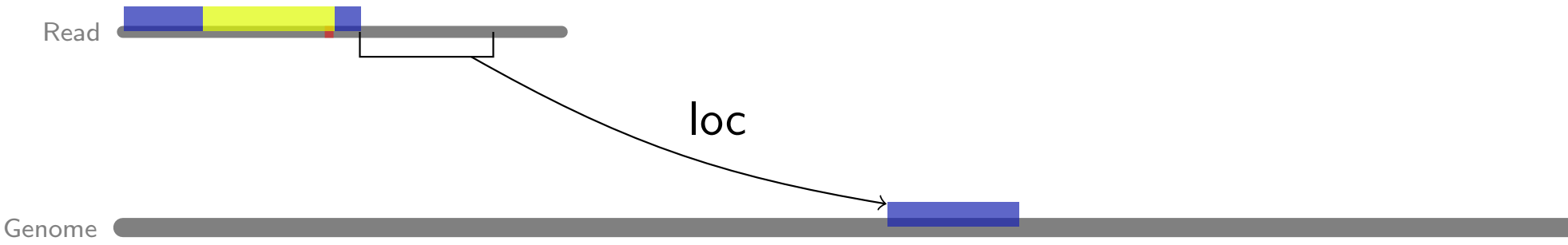
Using 22-mers to decipher a read



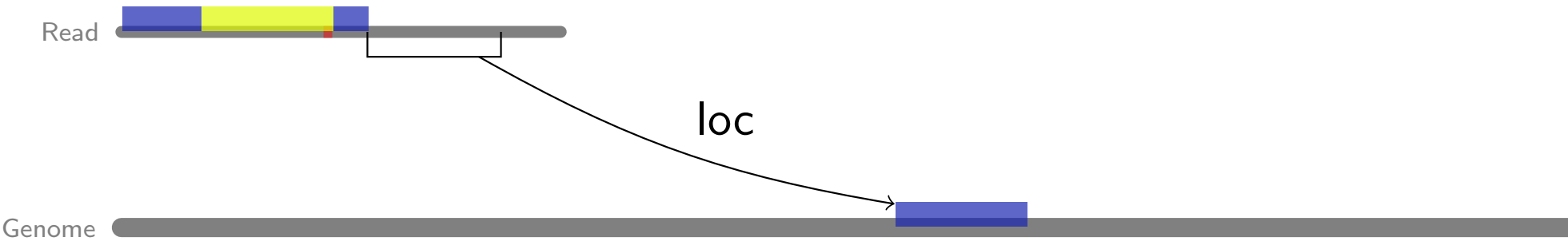
Using 22-mers to decipher a read



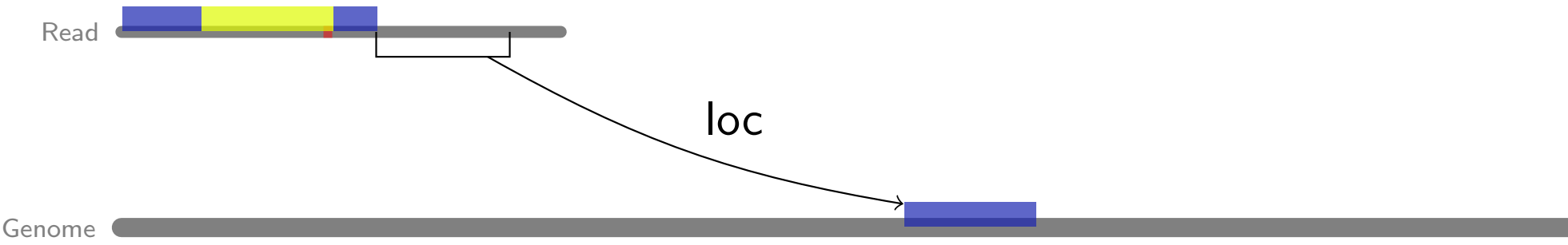
Using 22-mers to decipher a read



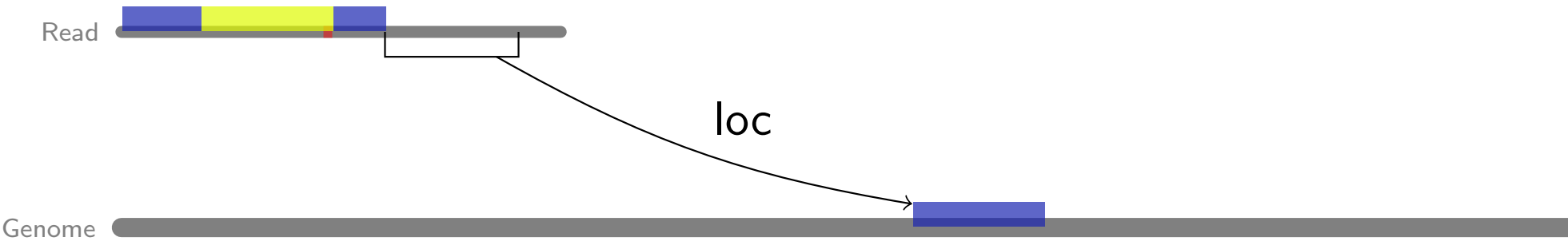
Using 22-mers to decipher a read



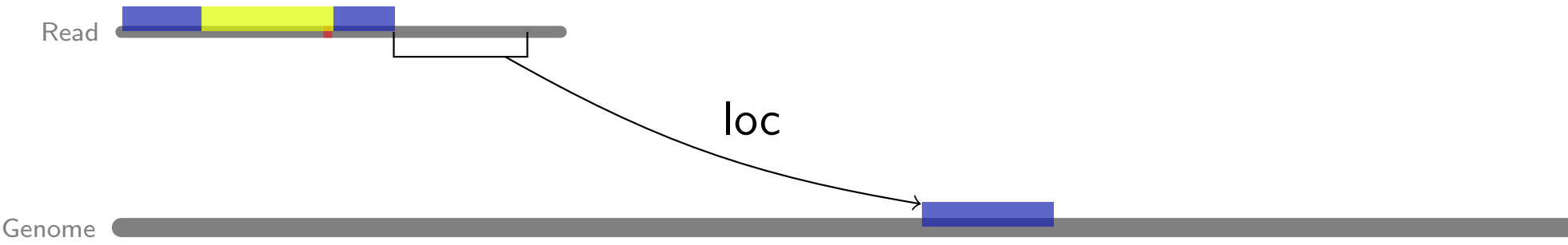
Using 22-mers to decipher a read



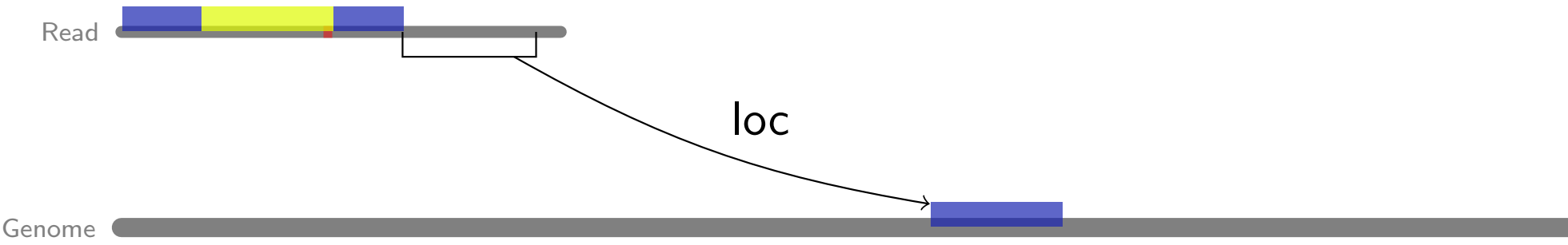
Using 22-mers to decipher a read



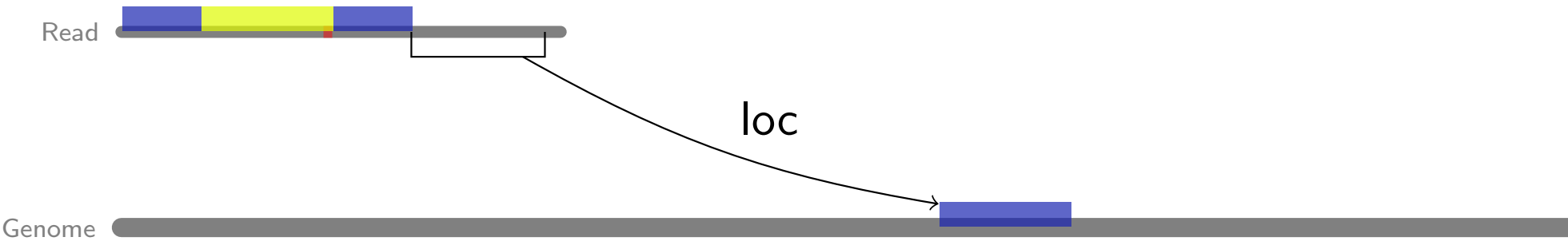
Using 22-mers to decipher a read



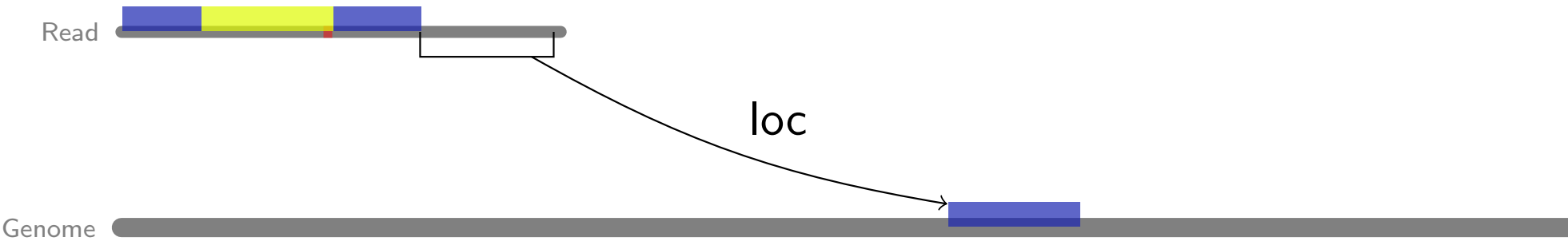
Using 22-mers to decipher a read



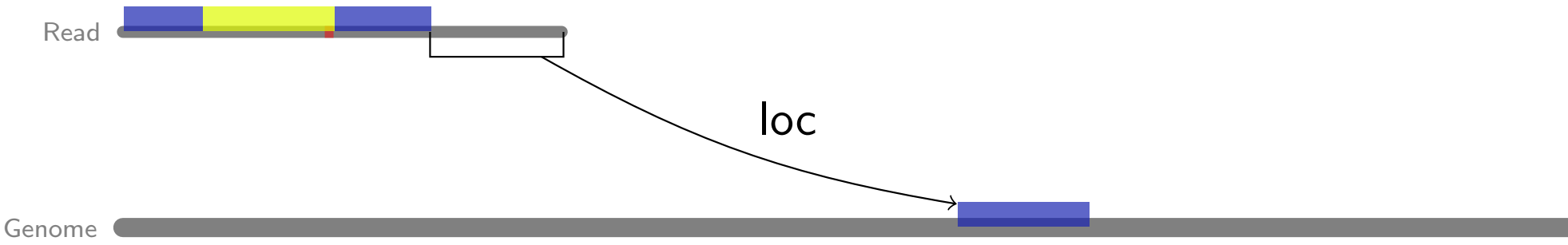
Using 22-mers to decipher a read



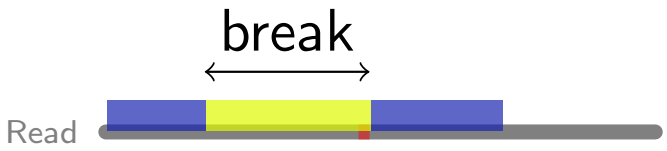
Using 22-mers to decipher a read



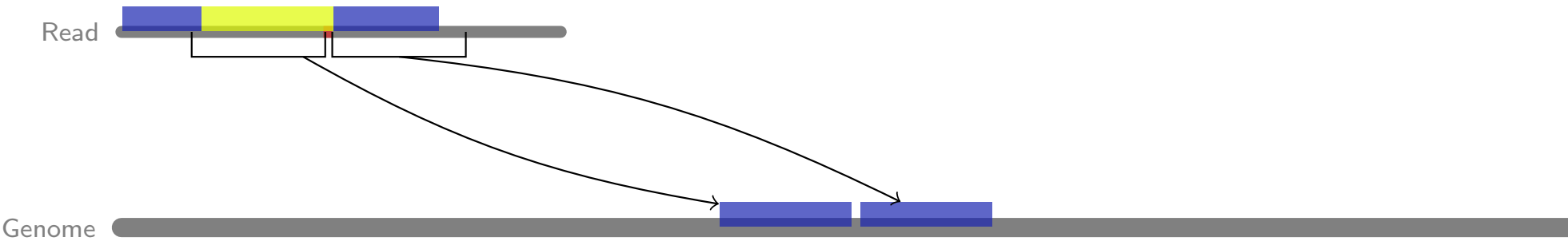
Using 22-mers to decipher a read



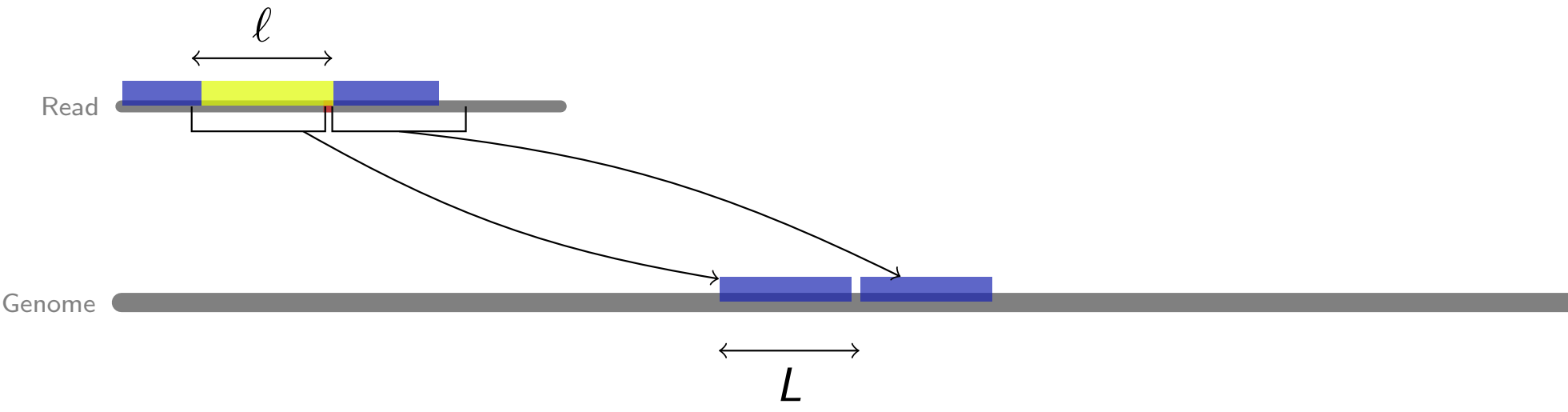
Using 22-mers to decipher a read



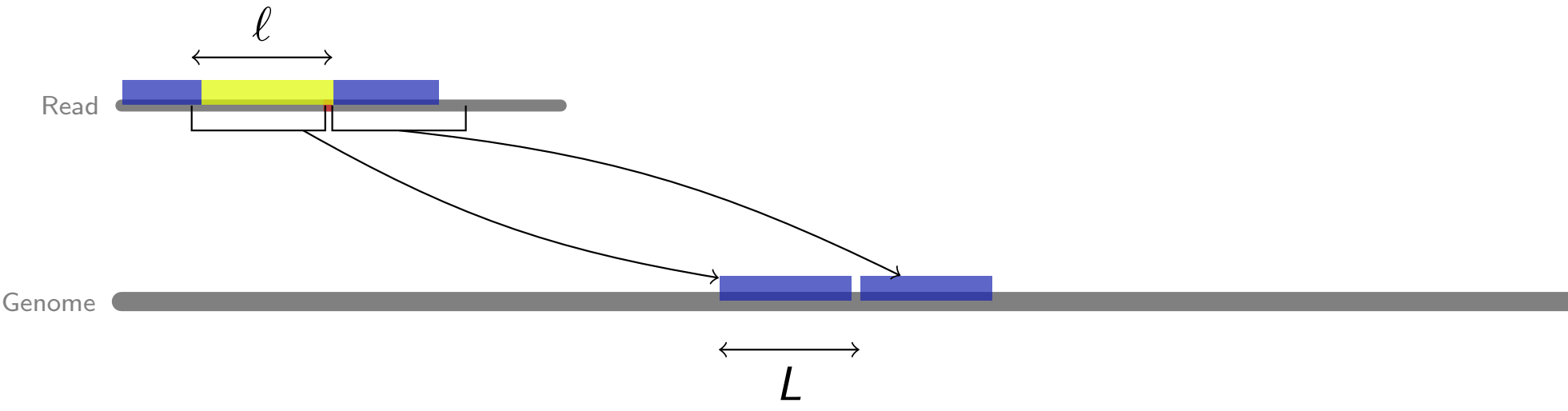
Using 22-mers to decipher a read



Using 22-mers to decipher a read

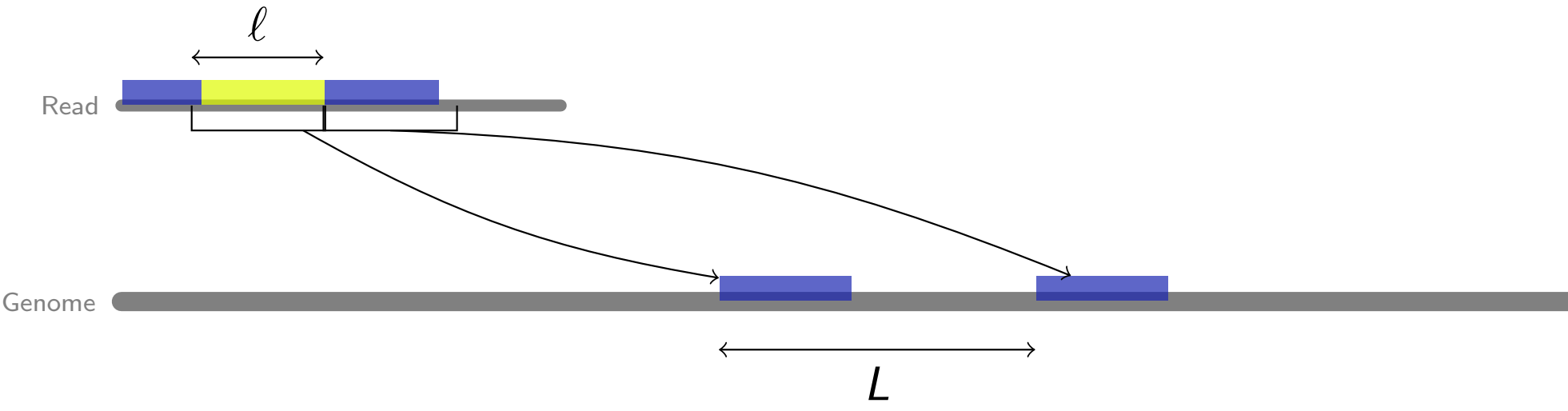


Using 22-mers to decipher a read



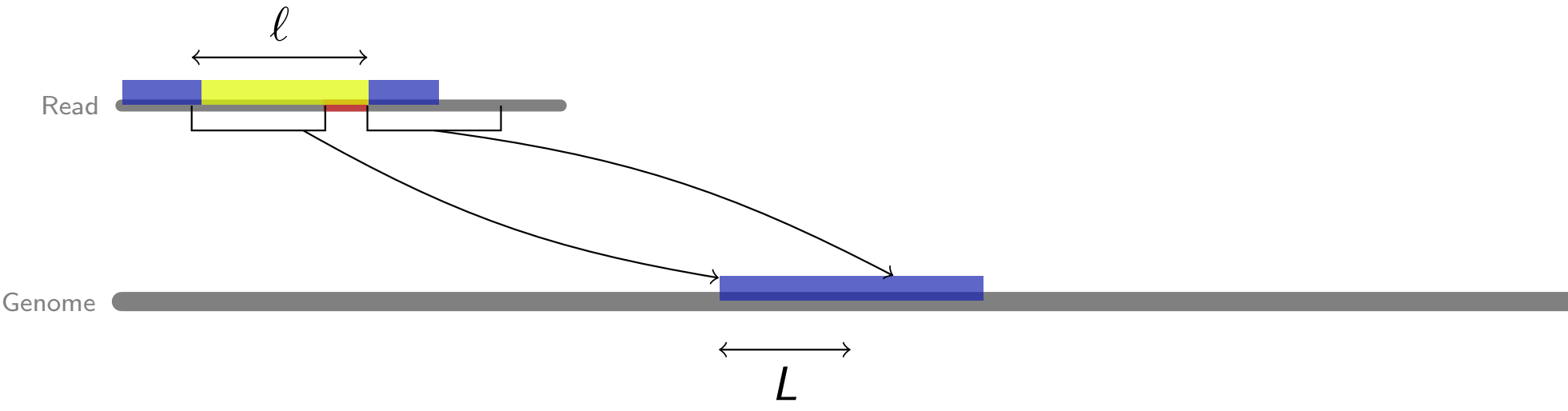
Substitution

Using 22-mers to decipher a read



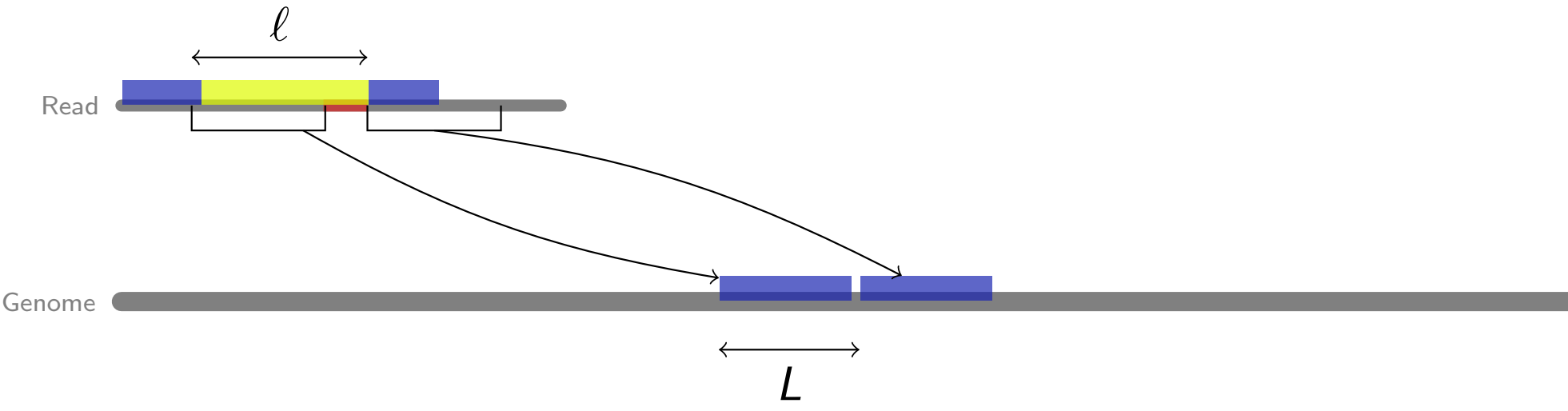
Deletion

Using 22-mers to decipher a read



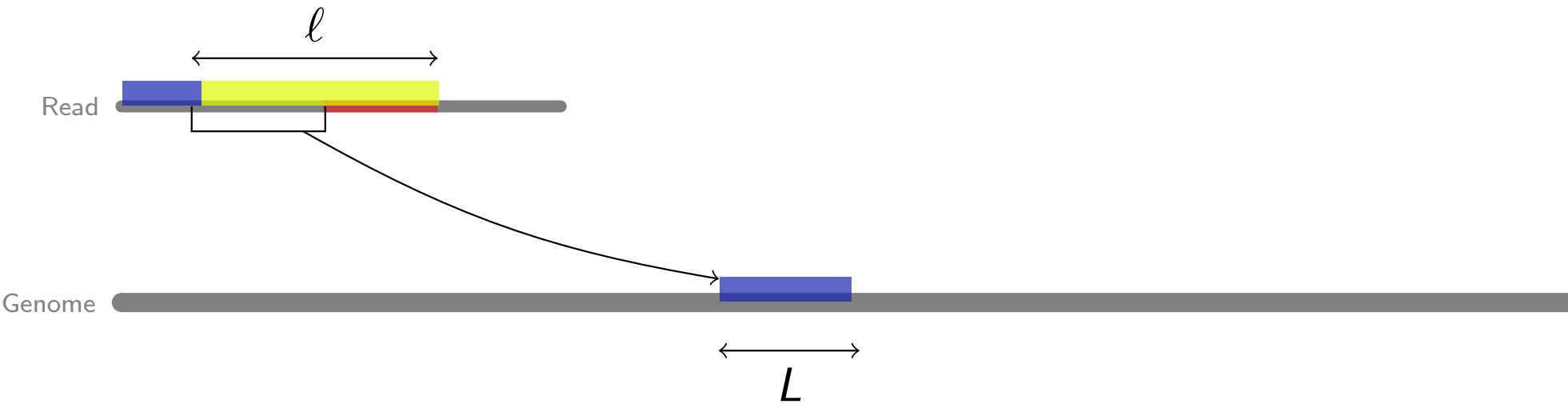
Insertion

Using 22-mers to decipher a read



$=$ $<$ $>$

Using 22-mers to decipher a read



We're not in an ideal world

Some vision problems. . .

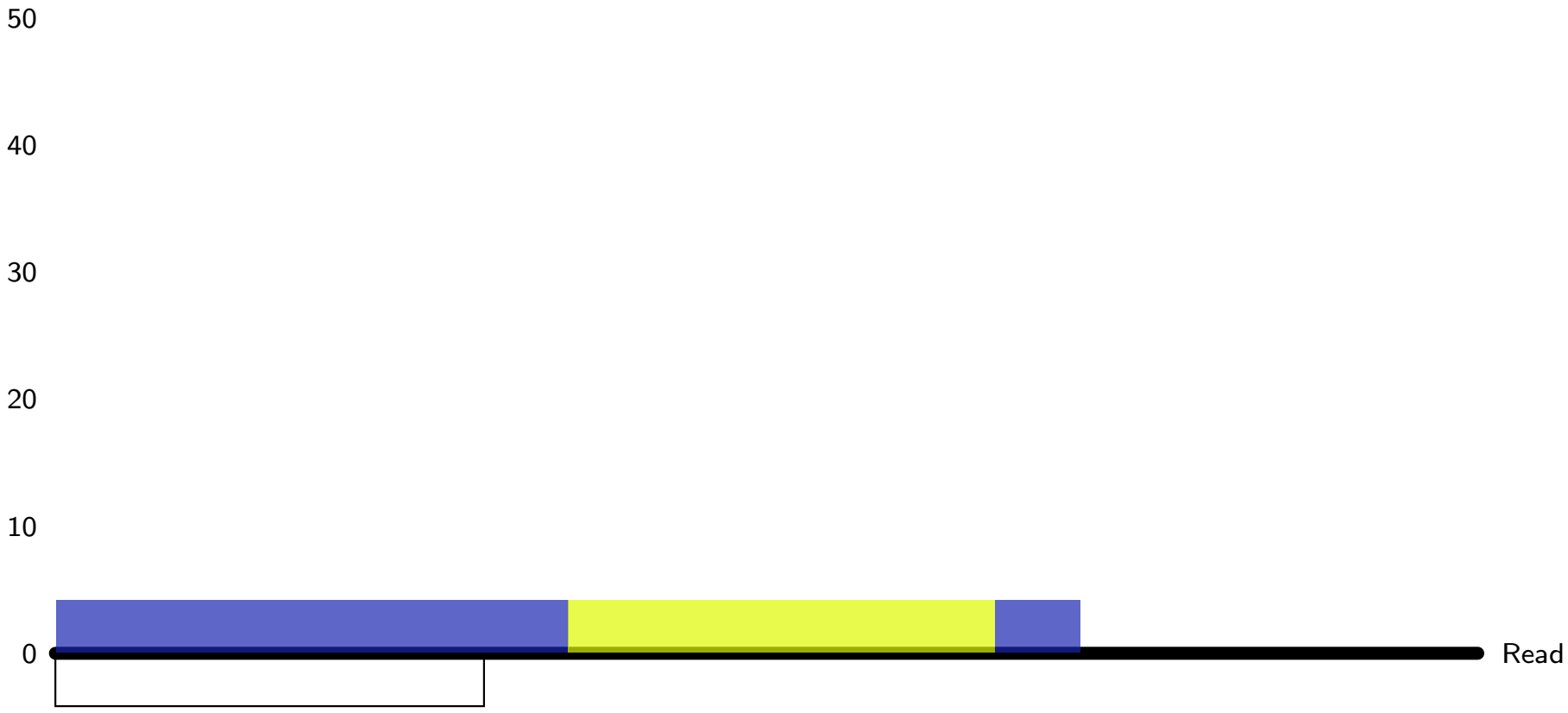
Error

or

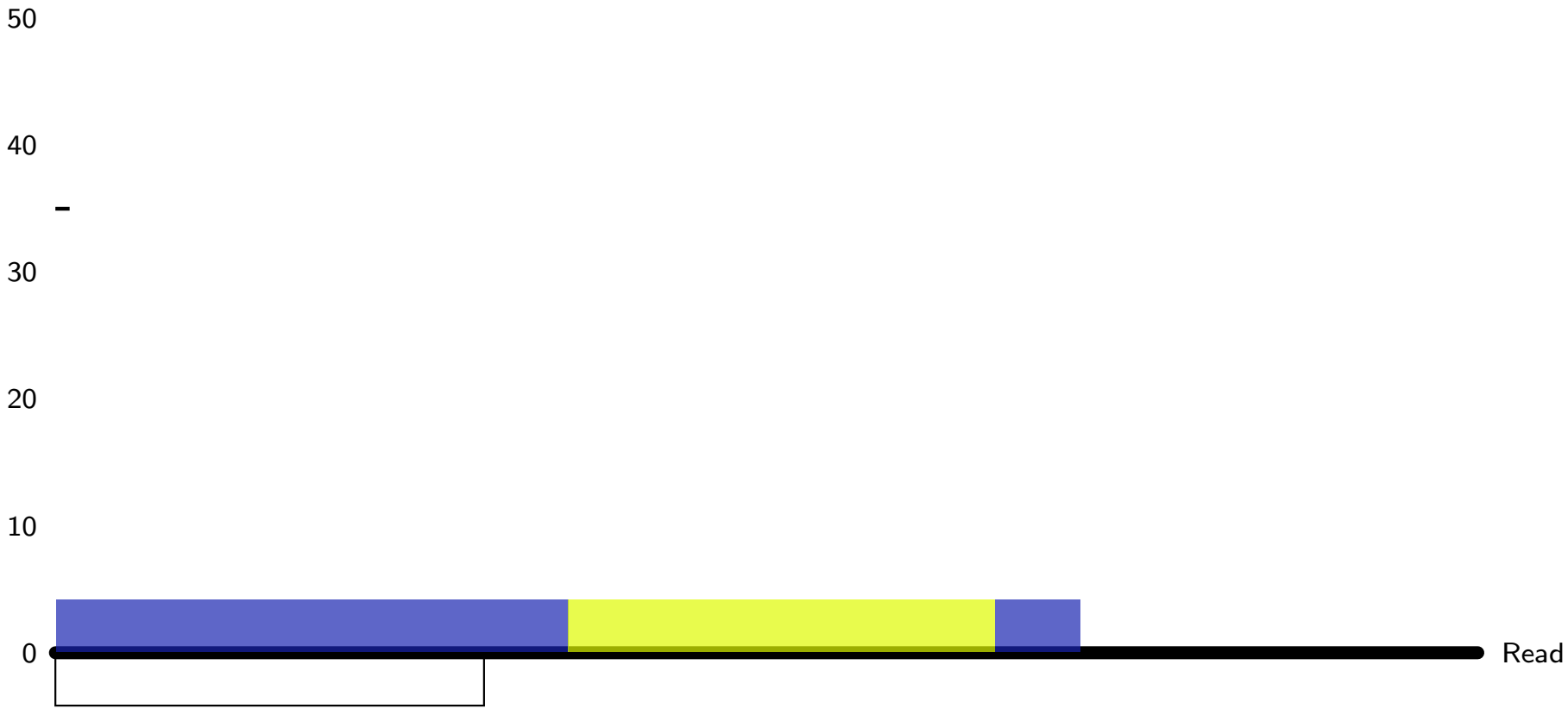
Variation?

Error or variation: using read collection

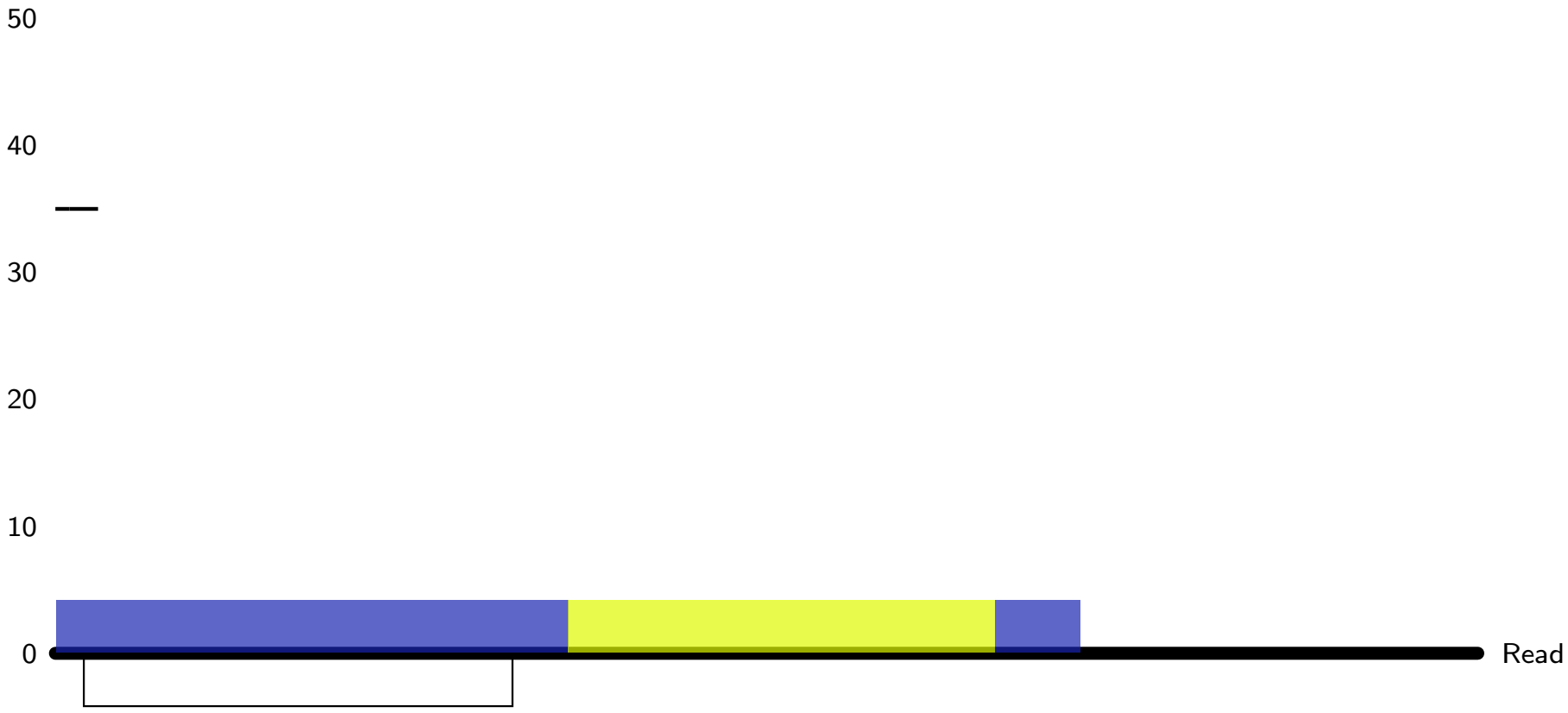
Error or variation: using read collection



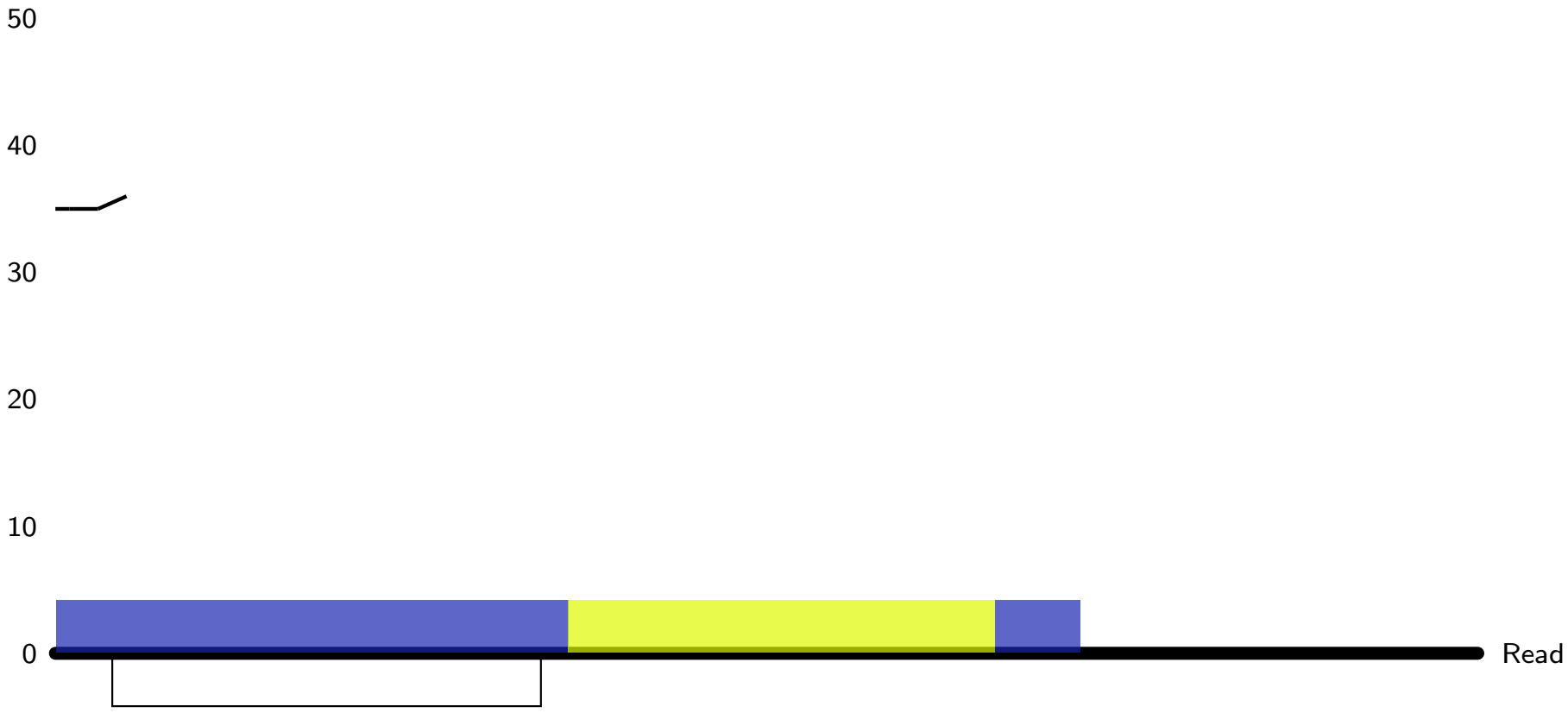
Error or variation: using read collection



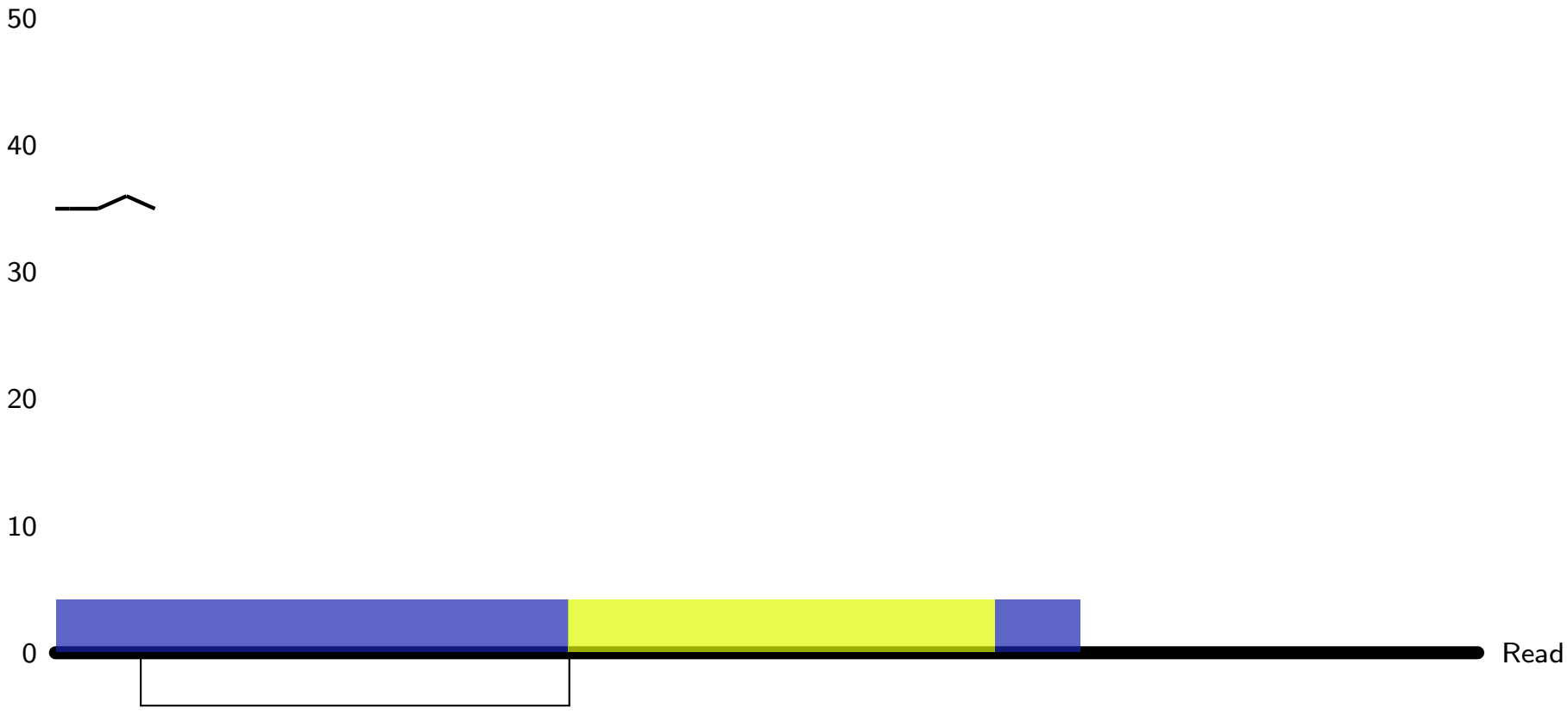
Error or variation: using read collection



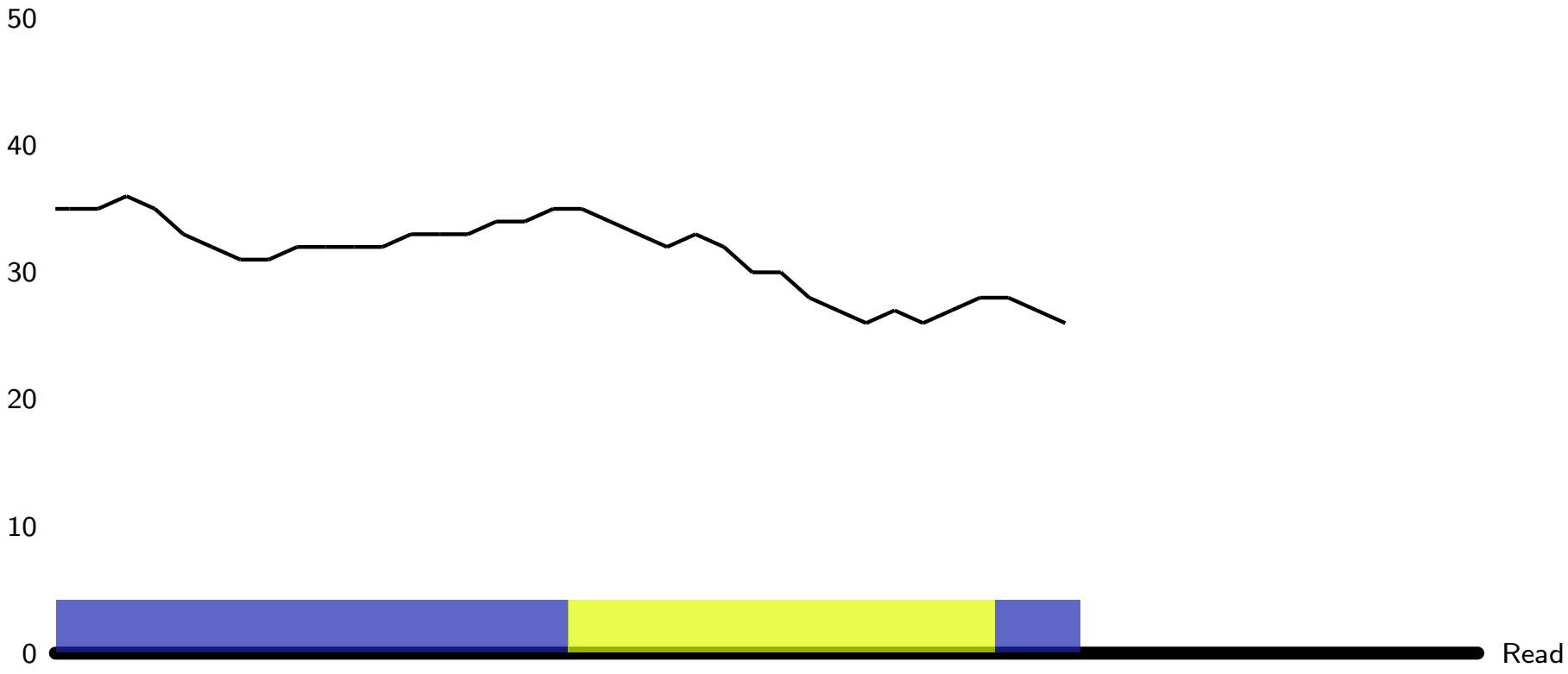
Error or variation: using read collection



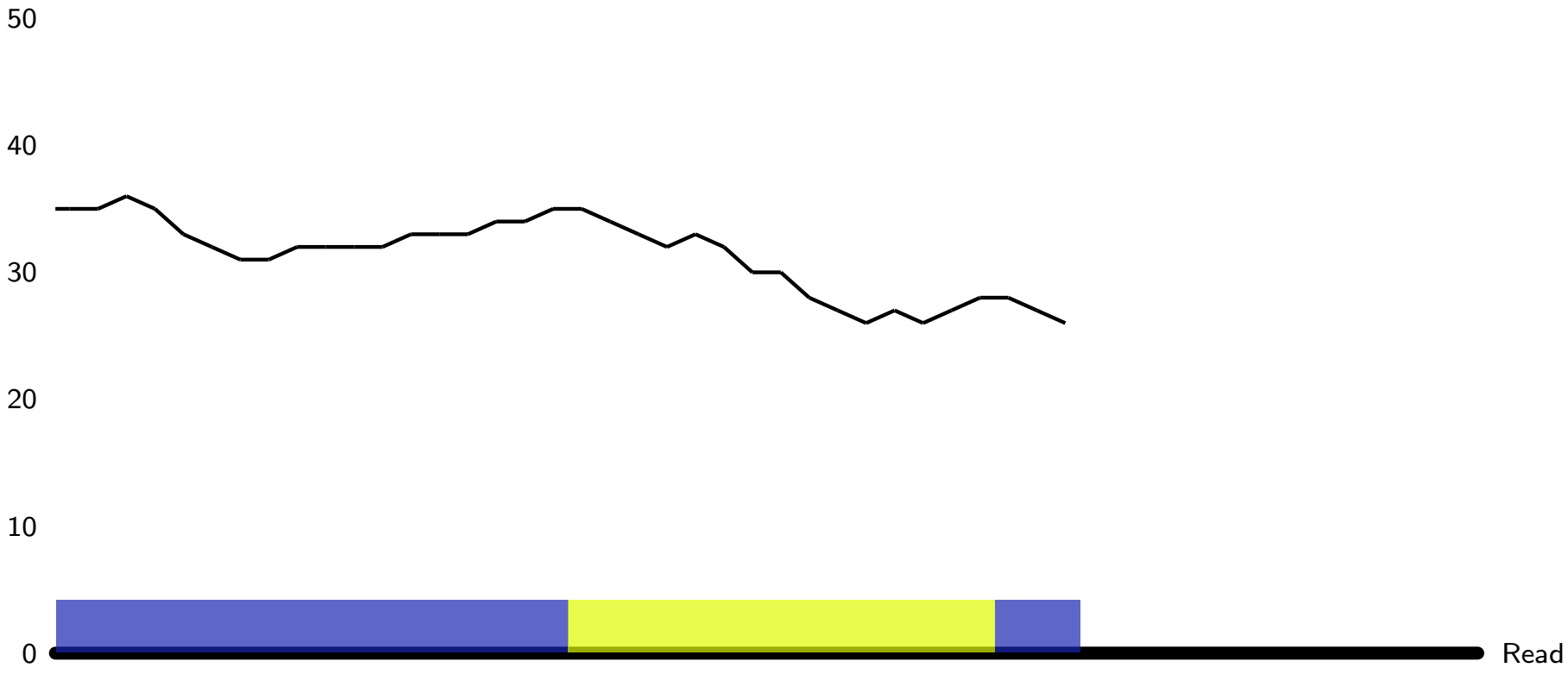
Error or variation: using read collection



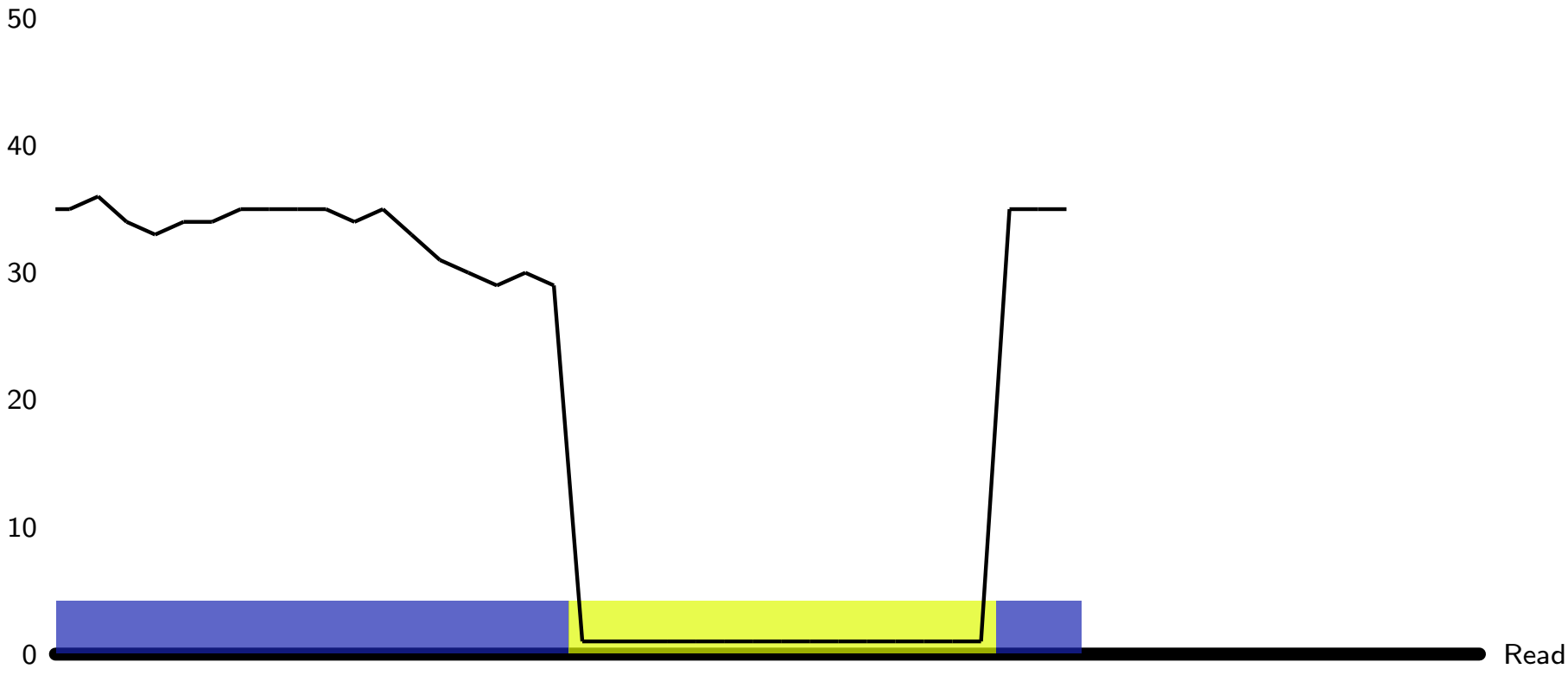
Error or variation: using read collection



Error or variation: using read collection



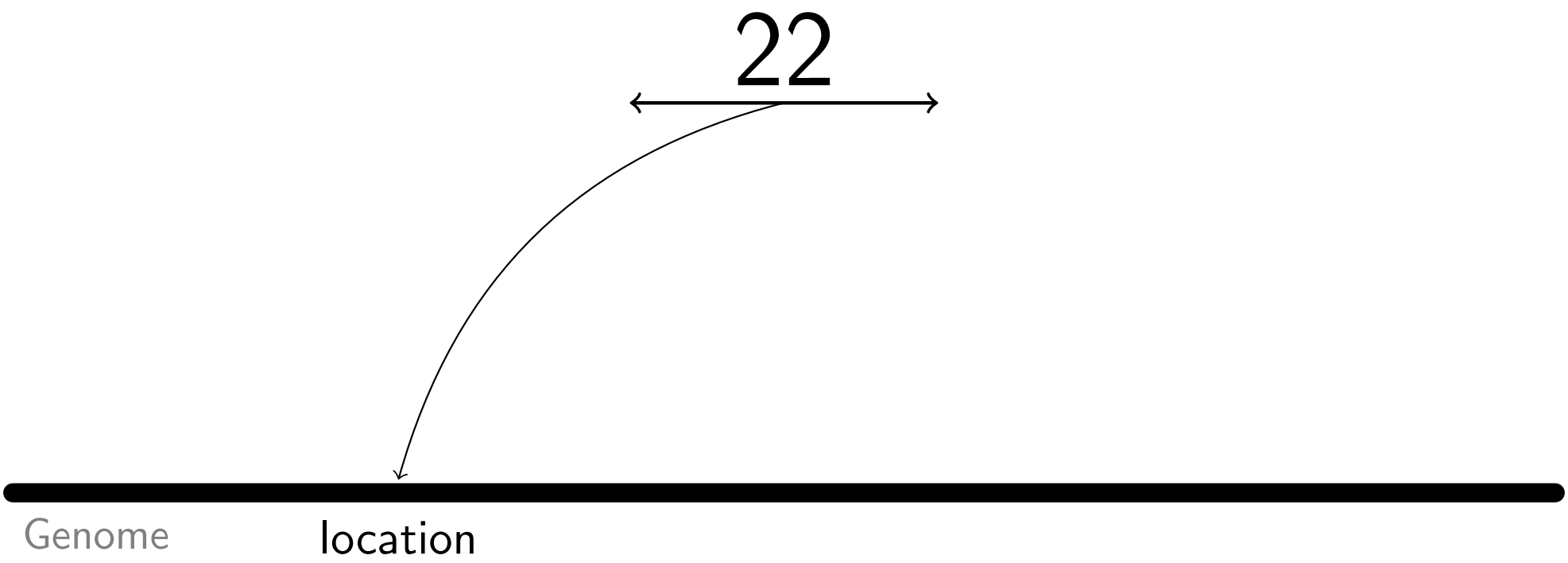
Error or variation: using read collection



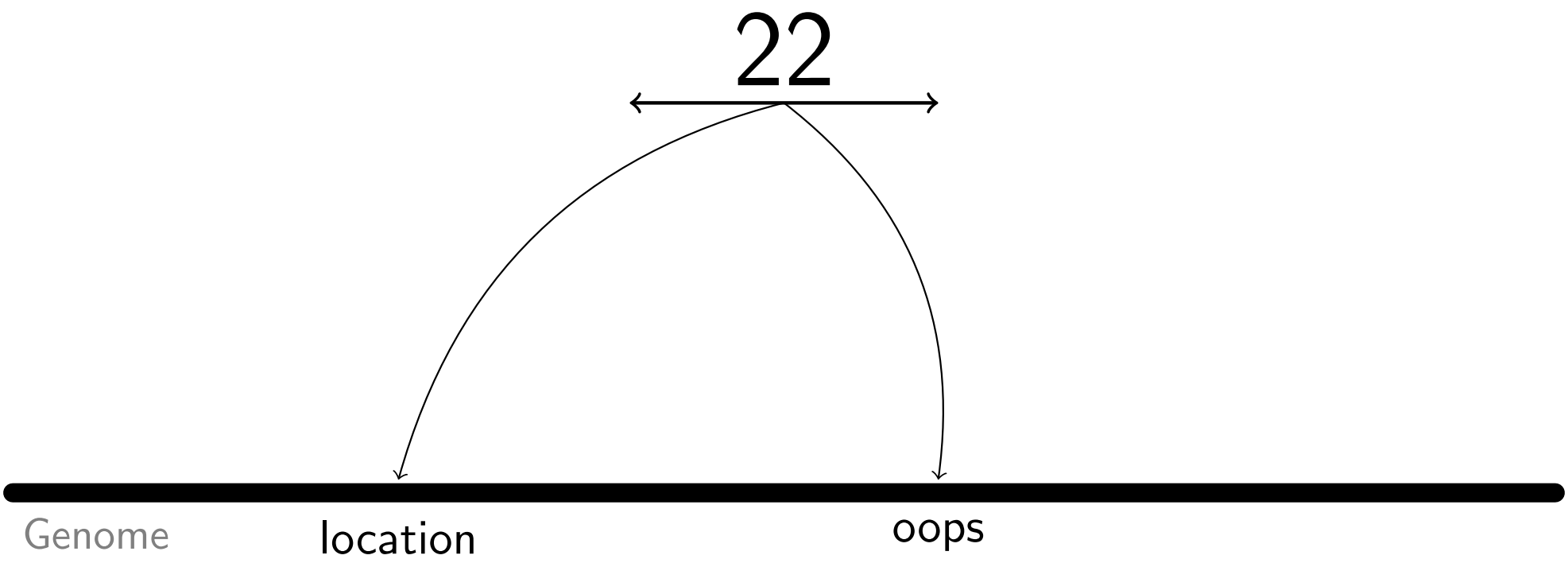
We're not in an ideal world

22, not such a magic number

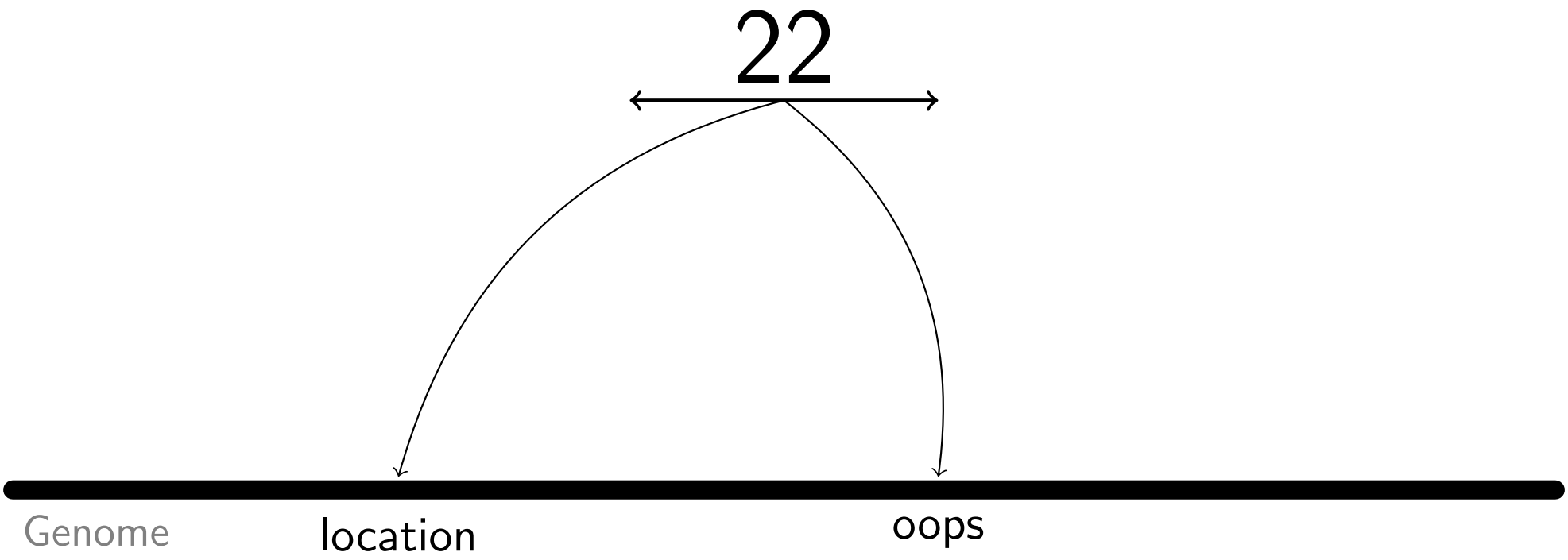
22, not such a magic number



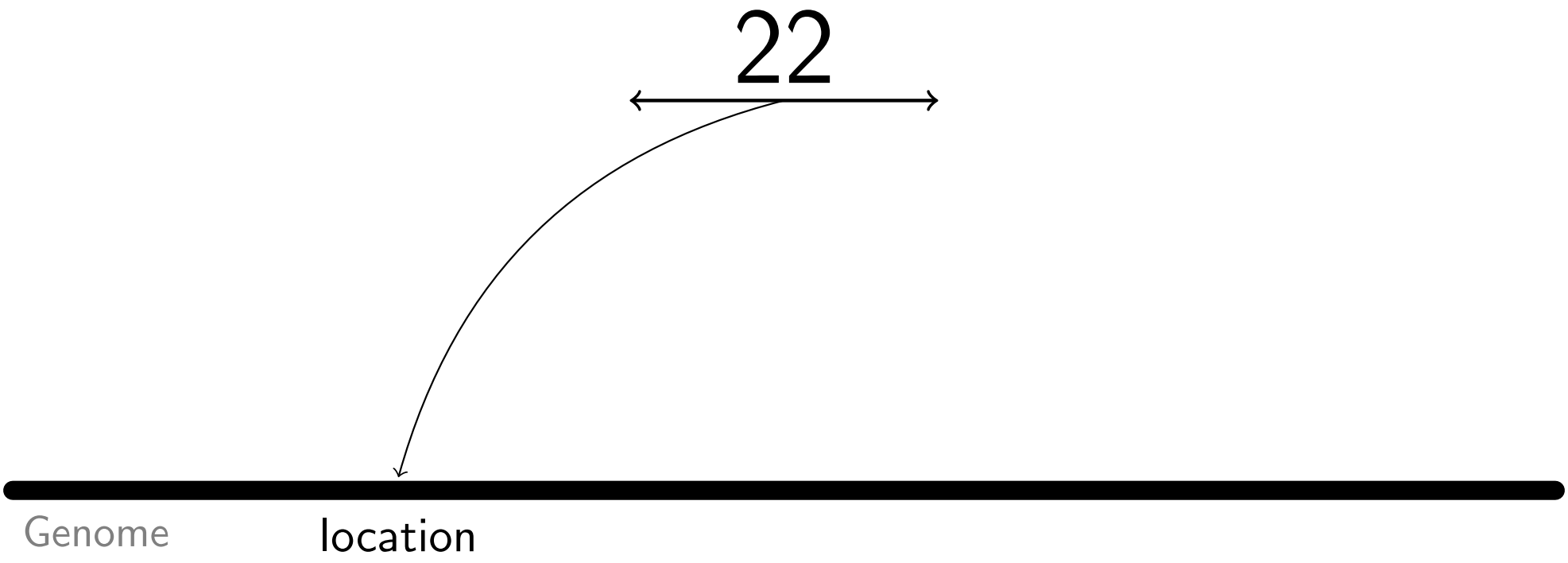
22, not such a magic number



22, not such a magic number



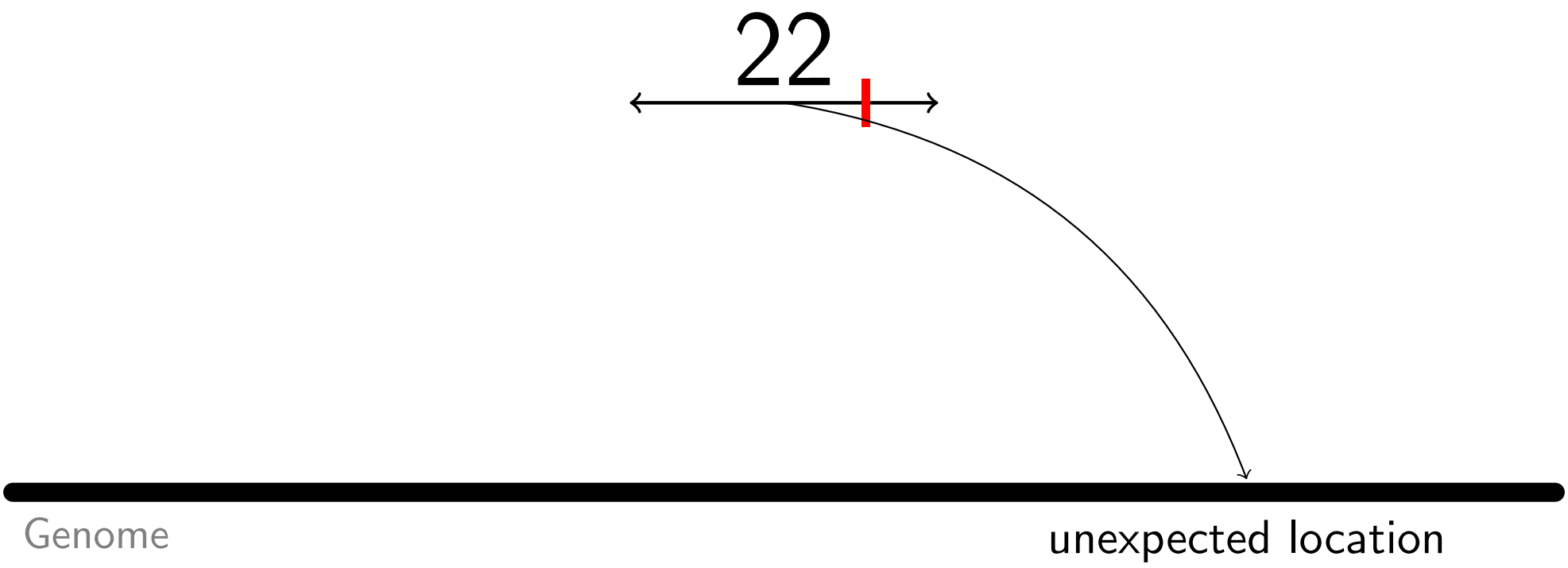
22, not such a magic number



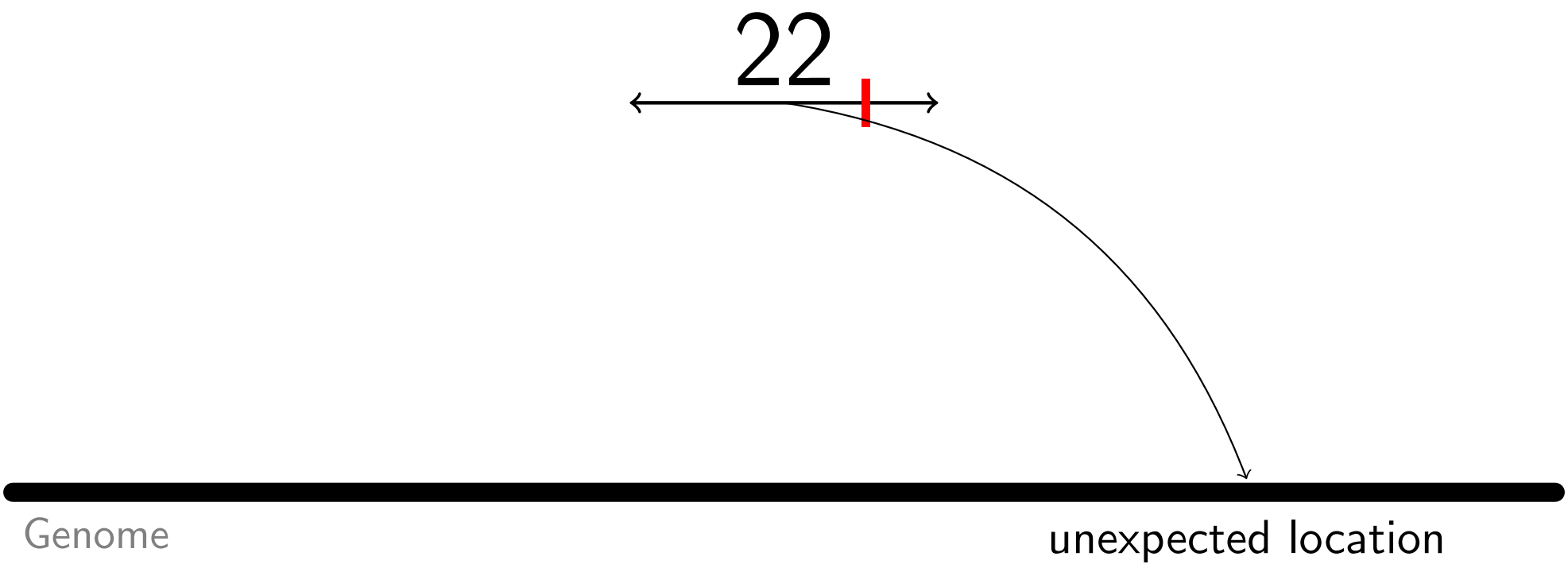
22, not such a magic number

Genome

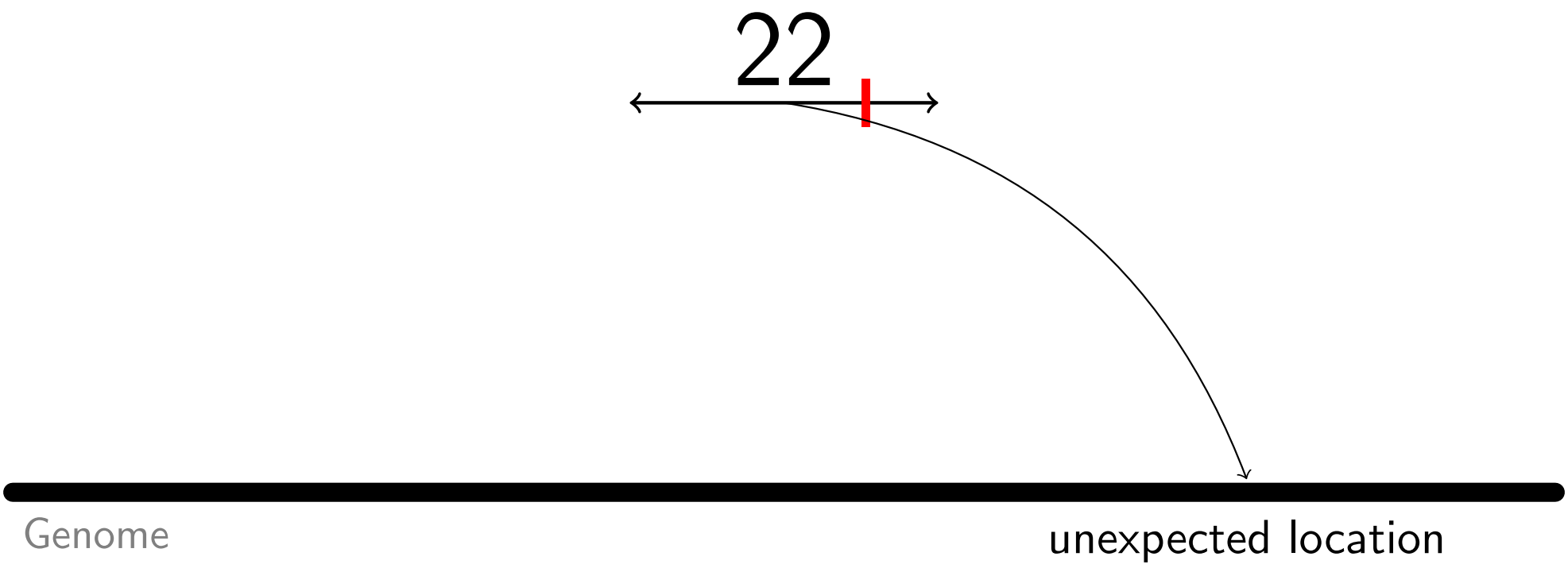
22, not such a magic number



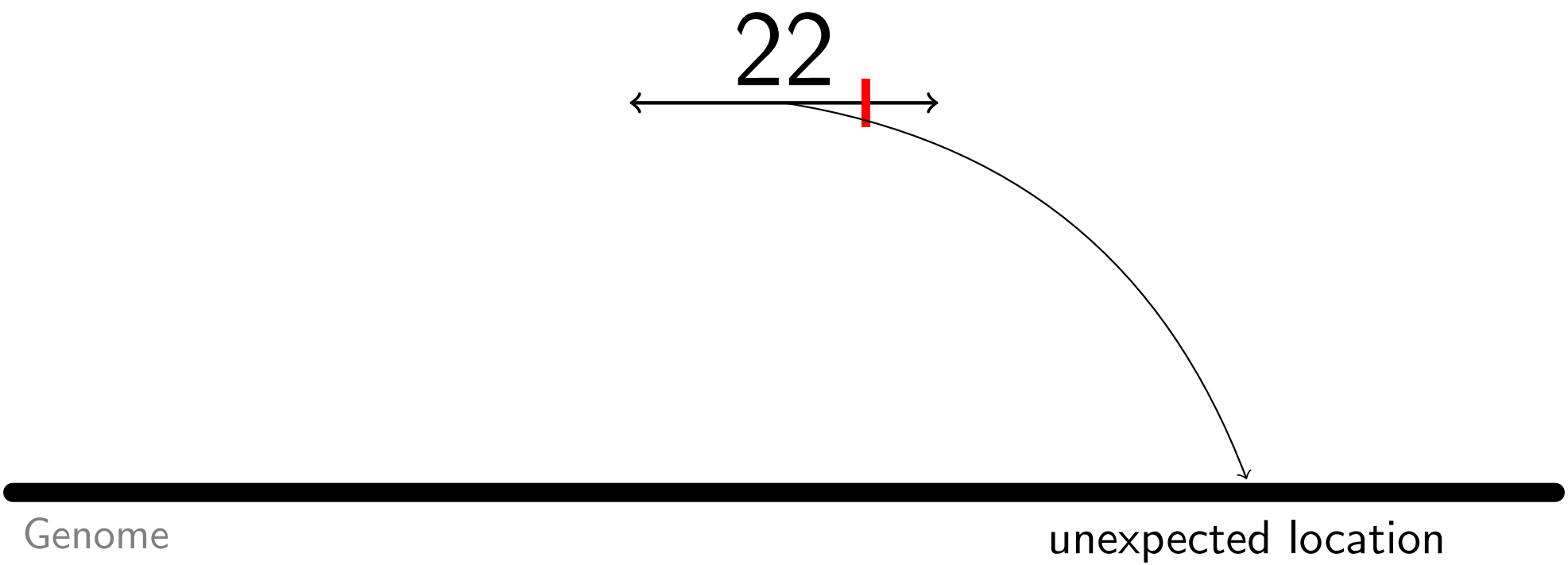
22, not such a magic number

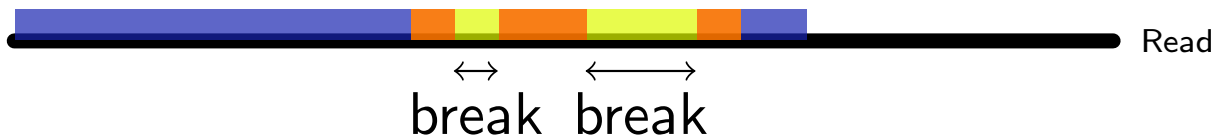


22, not such a magic number

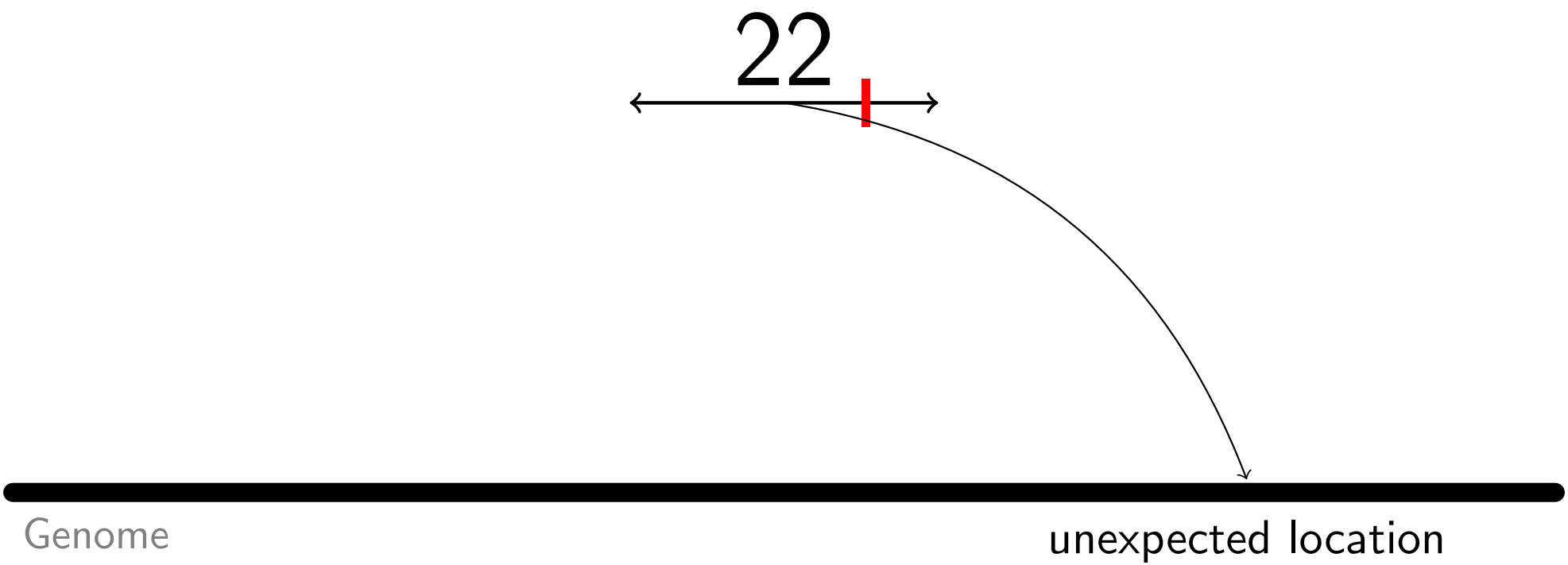


22, not such a magic number

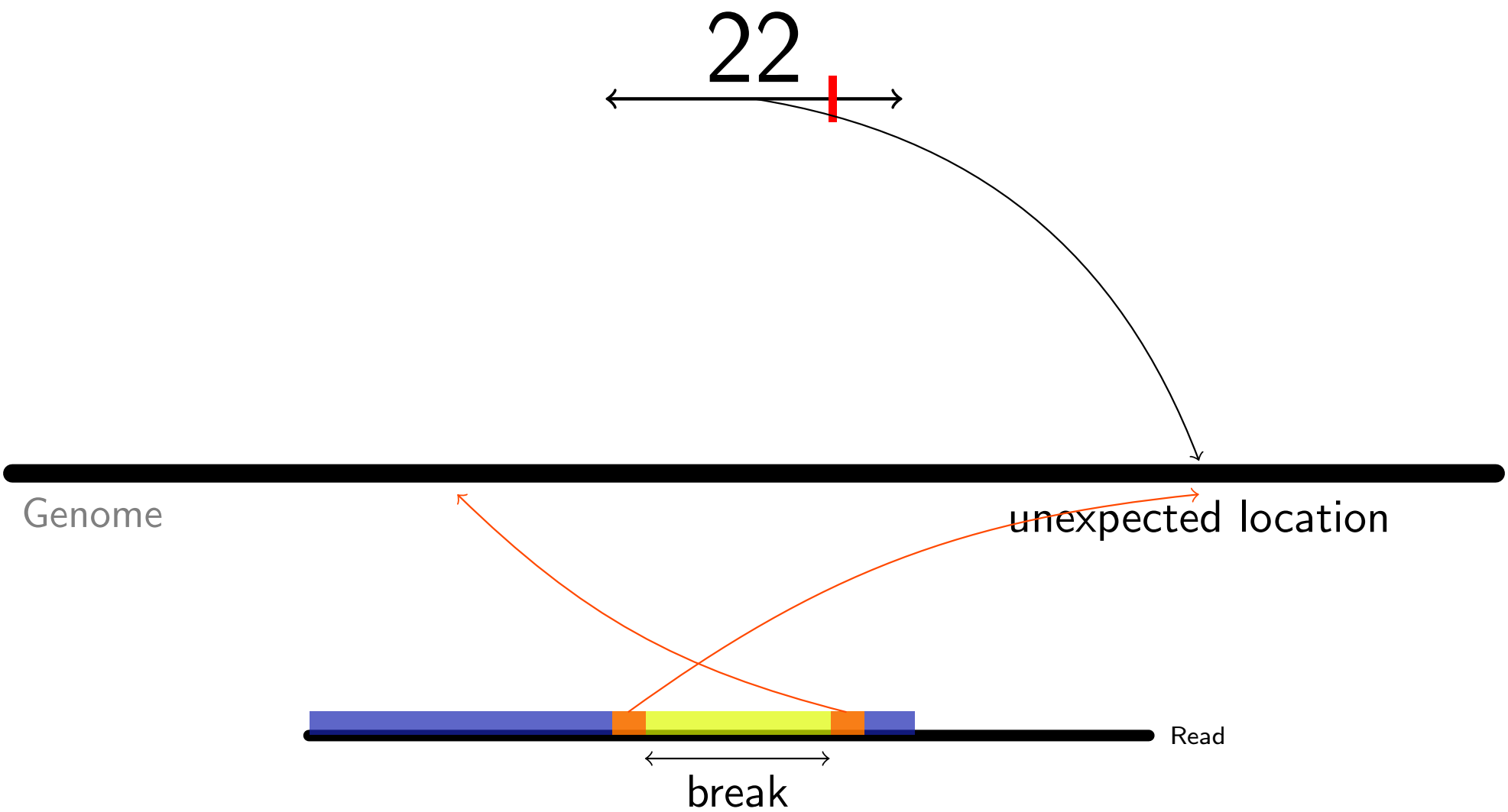




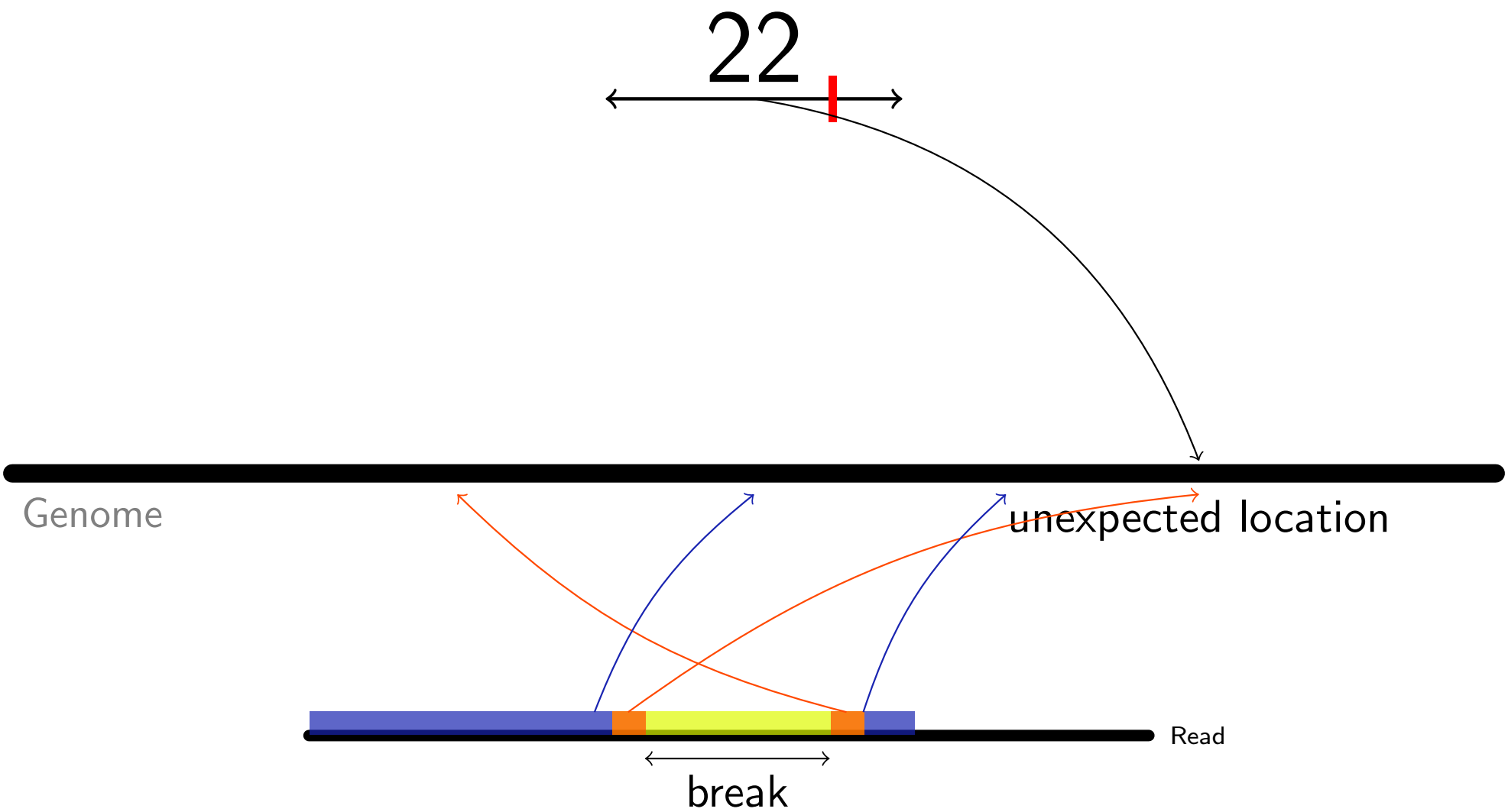
22, not such a magic number



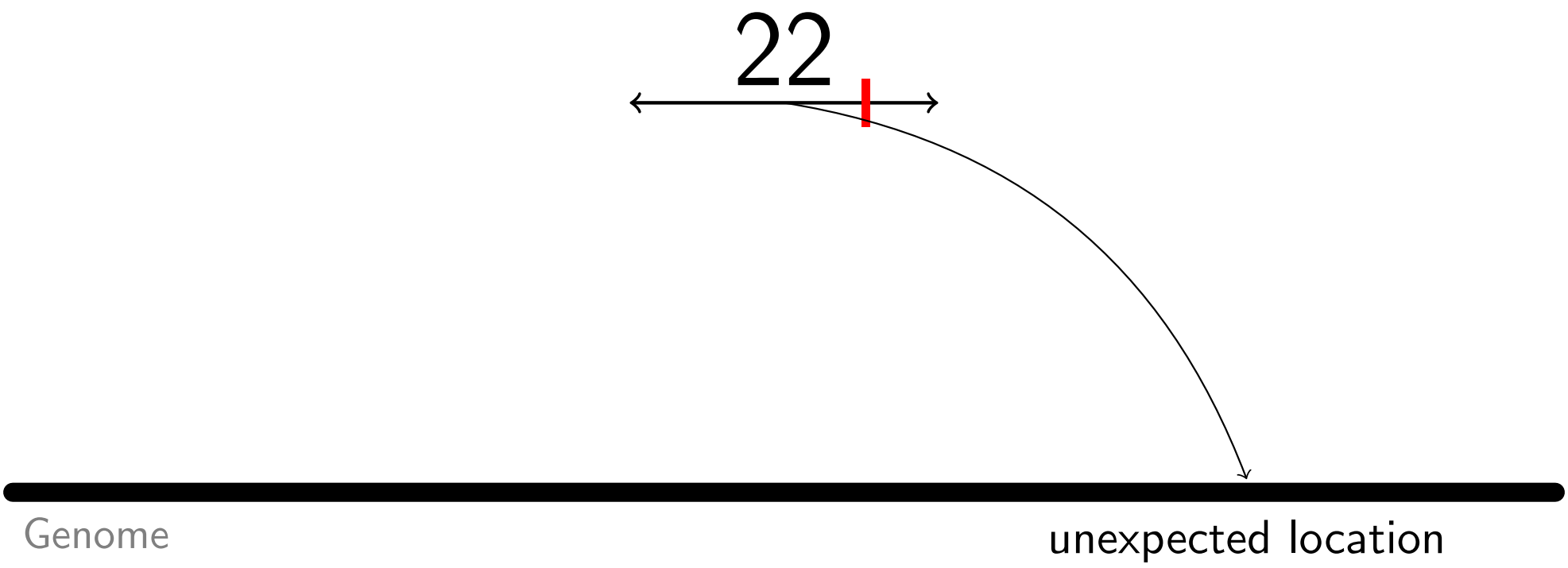
22, not such a magic number



22, not such a magic number



22, not such a magic number



CRAC: a global RNA-seq data analysis

CRAC: a global RNA-seq data analysis

Read mapping

CRAC: a global RNA-seq data analysis

Read mapping

Substitutions

CRAC: a global RNA-seq data analysis

Read mapping

Substitutions

Short indels

CRAC: a global RNA-seq data analysis

Read mapping

Substitutions

Short indels

Splicing events

CRAC: a global RNA-seq data analysis

Read mapping

Substitutions

Short indels

Splicing events

Fusions

CRAC: a global RNA-seq data analysis

Read mapping

Substitutions

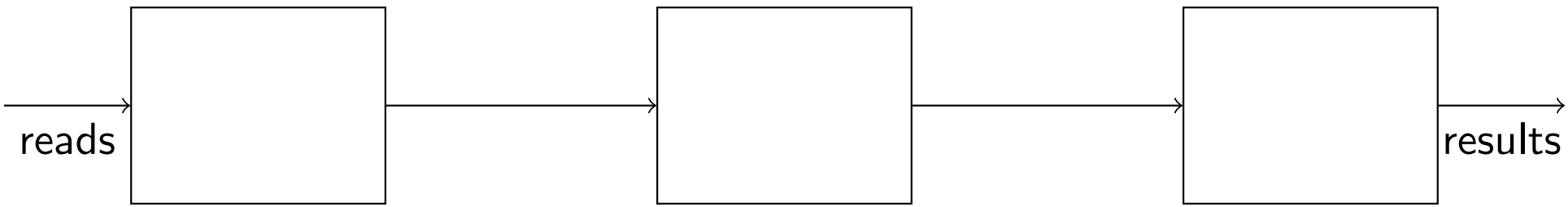
Short indels

Splicing events

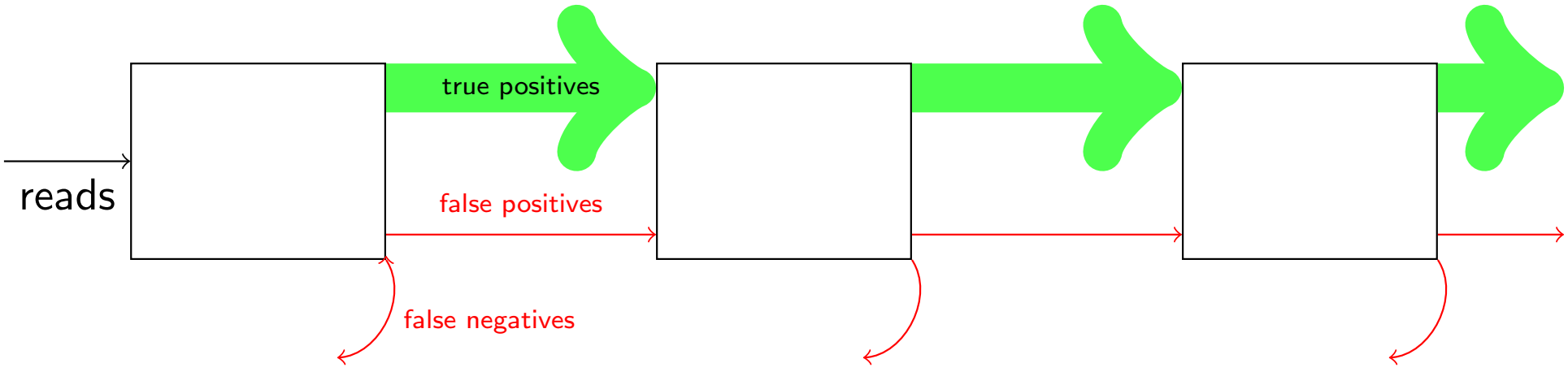
Fusions

(Sequencing errors)

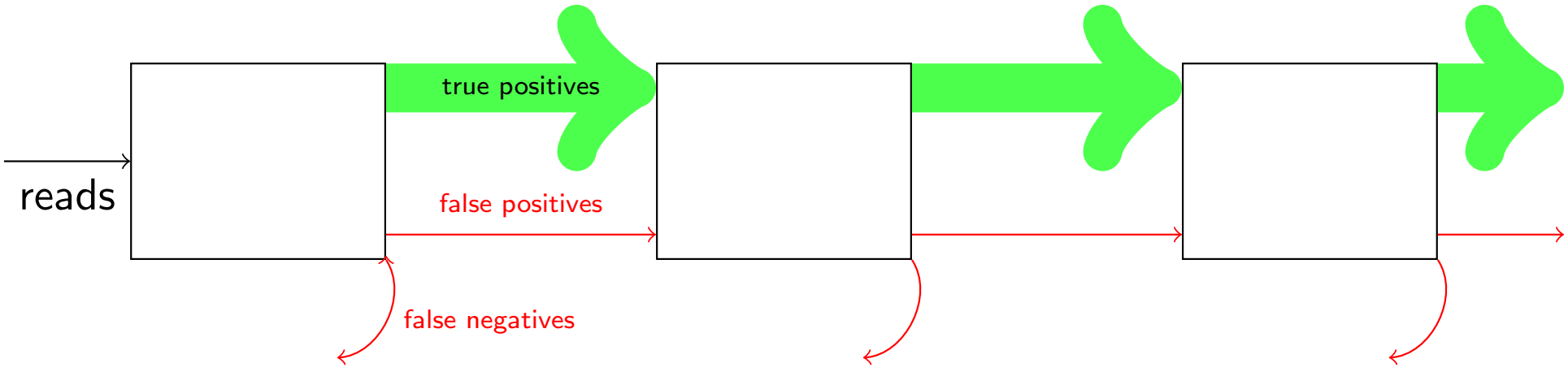
Keep it simple and stupid?



Keep it simple and stupid?



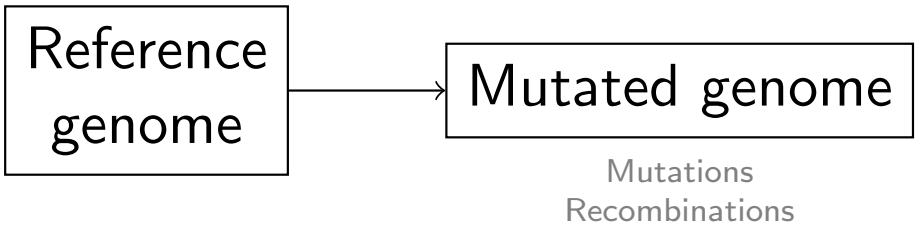
Keep it simple and stupid?



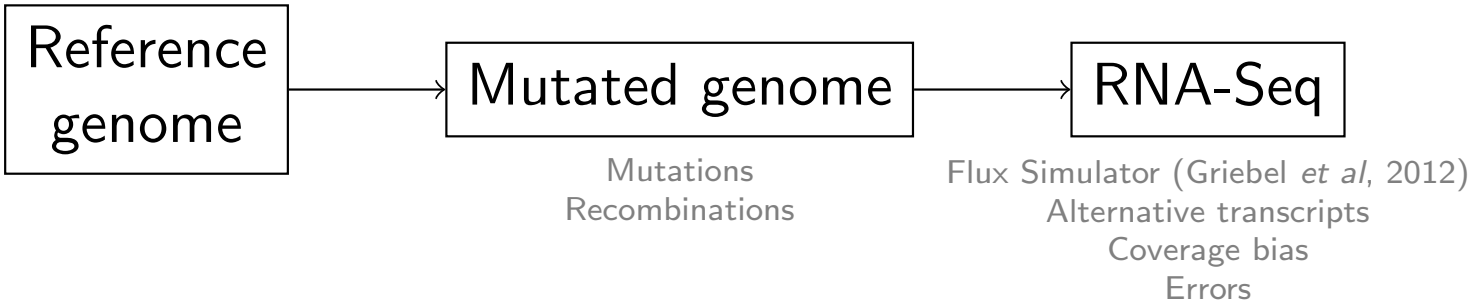
One tool may be better than many. . .

Simulated RNA-Seq data

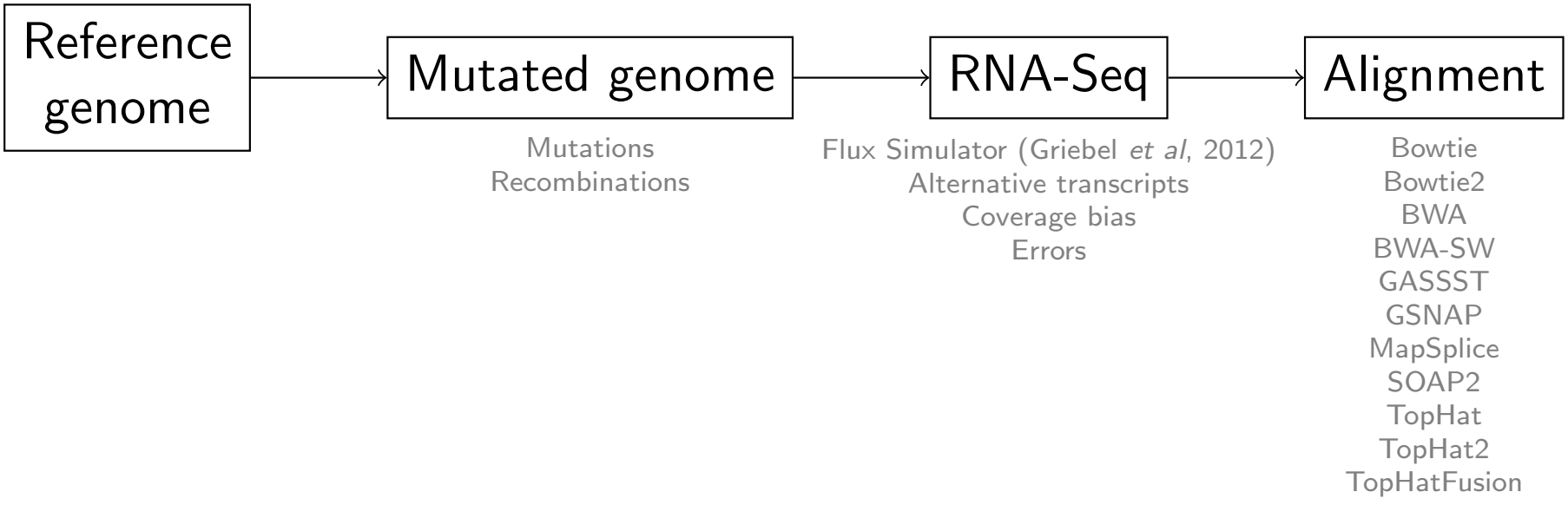
Simulated RNA-Seq data



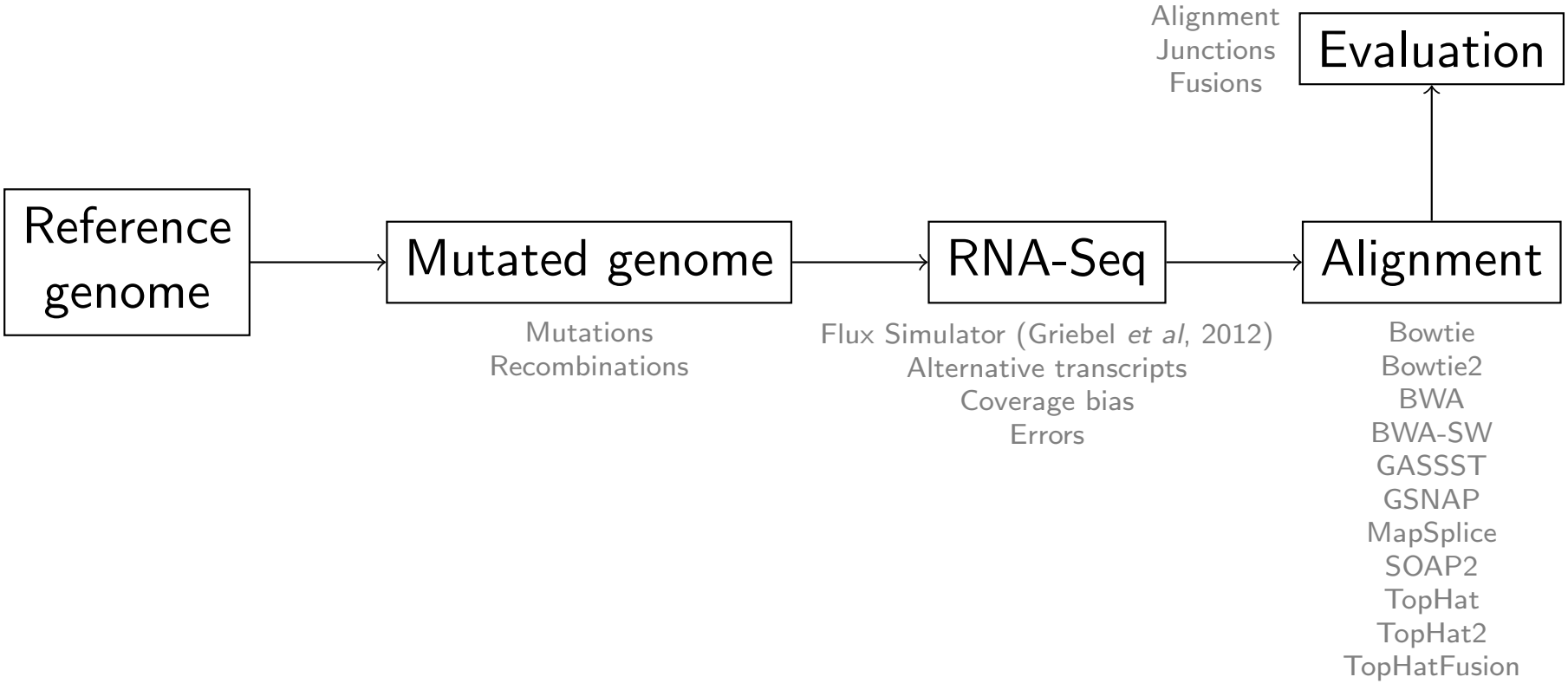
Simulated RNA-Seq data



Simulated RNA-Seq data



Simulated RNA-Seq data



Datasets

45M reads

75bp

29k SNVs
2.5k insertions
2.5k deletions
647 fusions
13M errors

Datasets

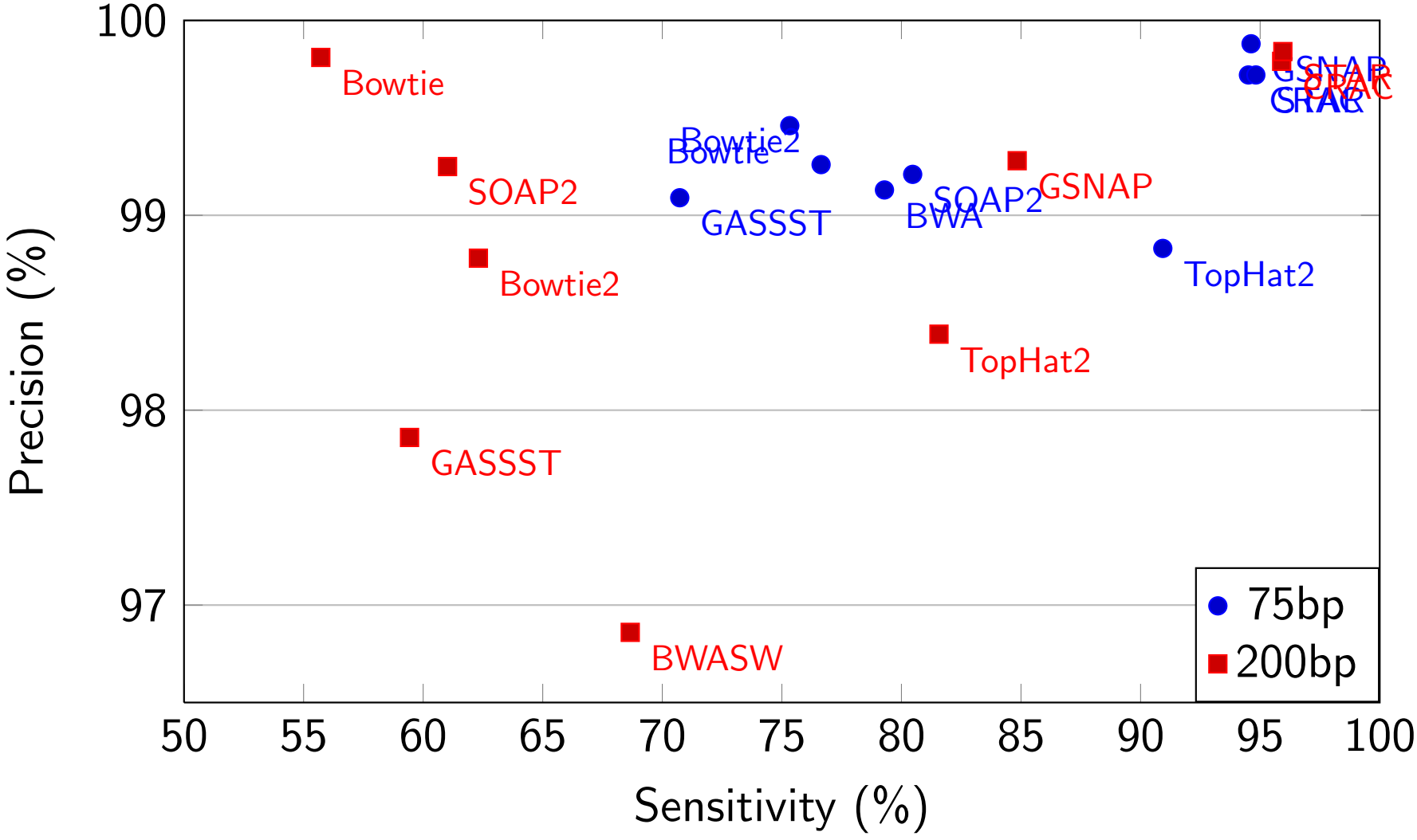
45M reads 75bp

29k SNVs
2.5k insertions
2.5k deletions
647 fusions
13M errors

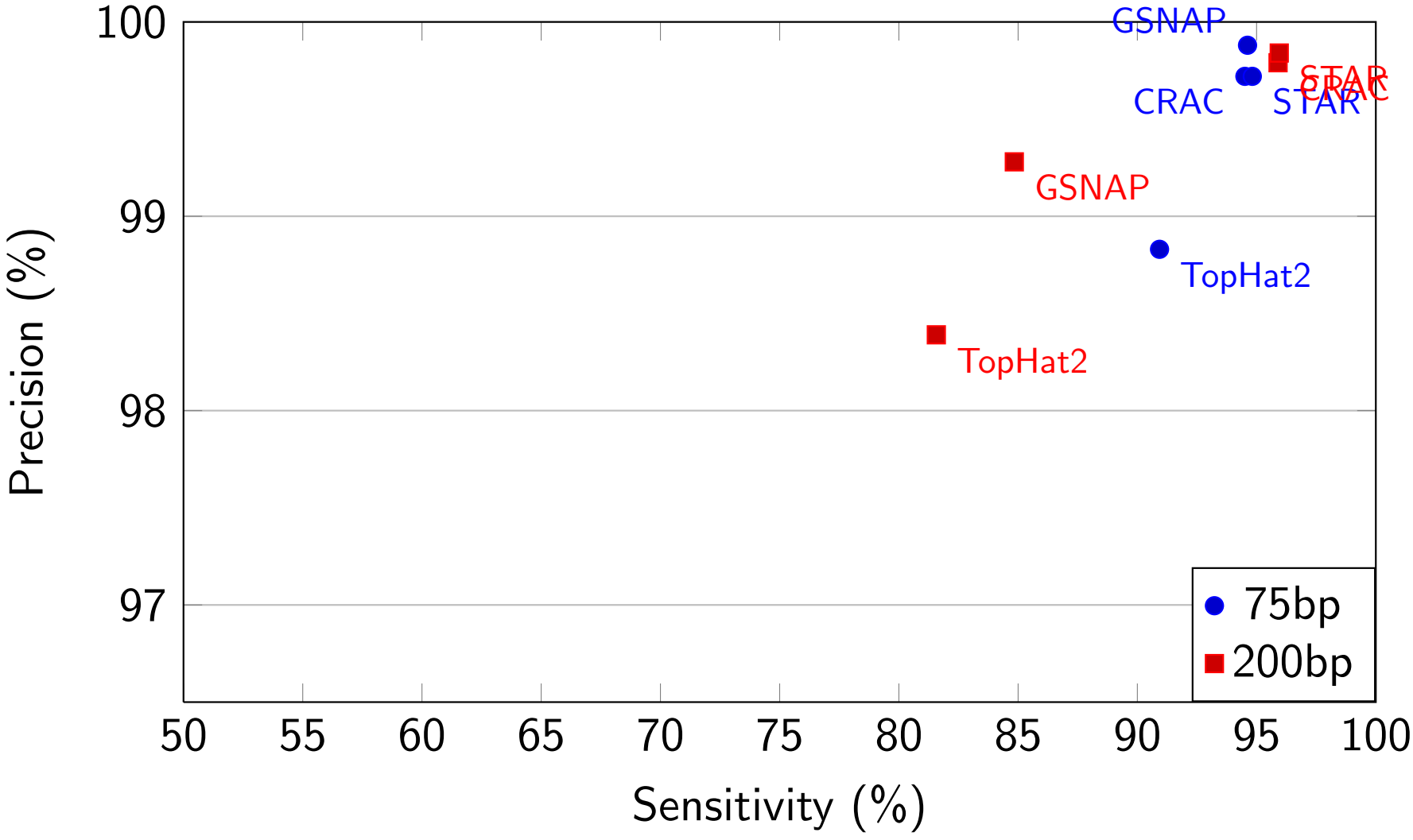
48M reads 200bp

53k SNVs
5k insertions
5k deletions
914 fusions
39M errors

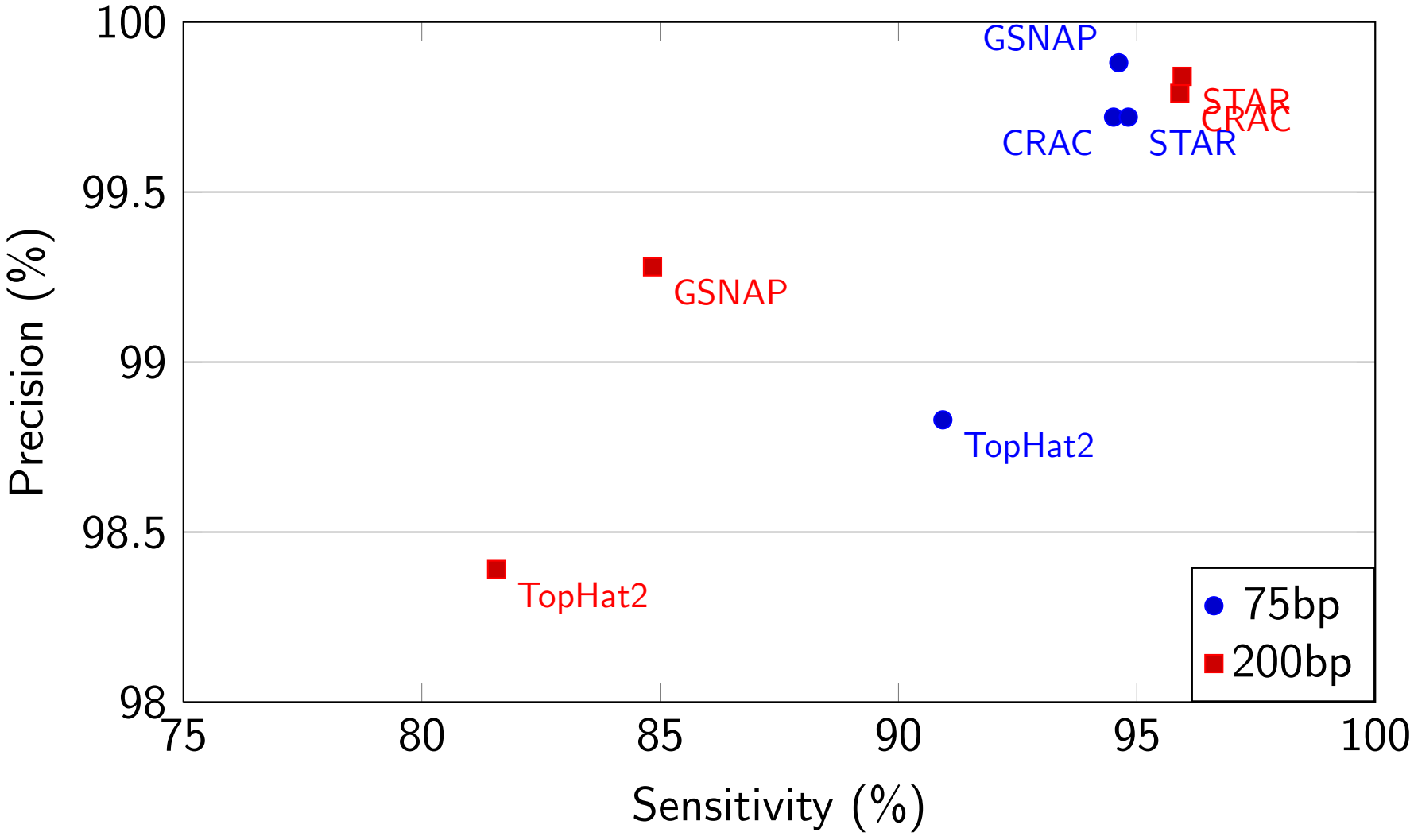
Alignment (single occurrences)



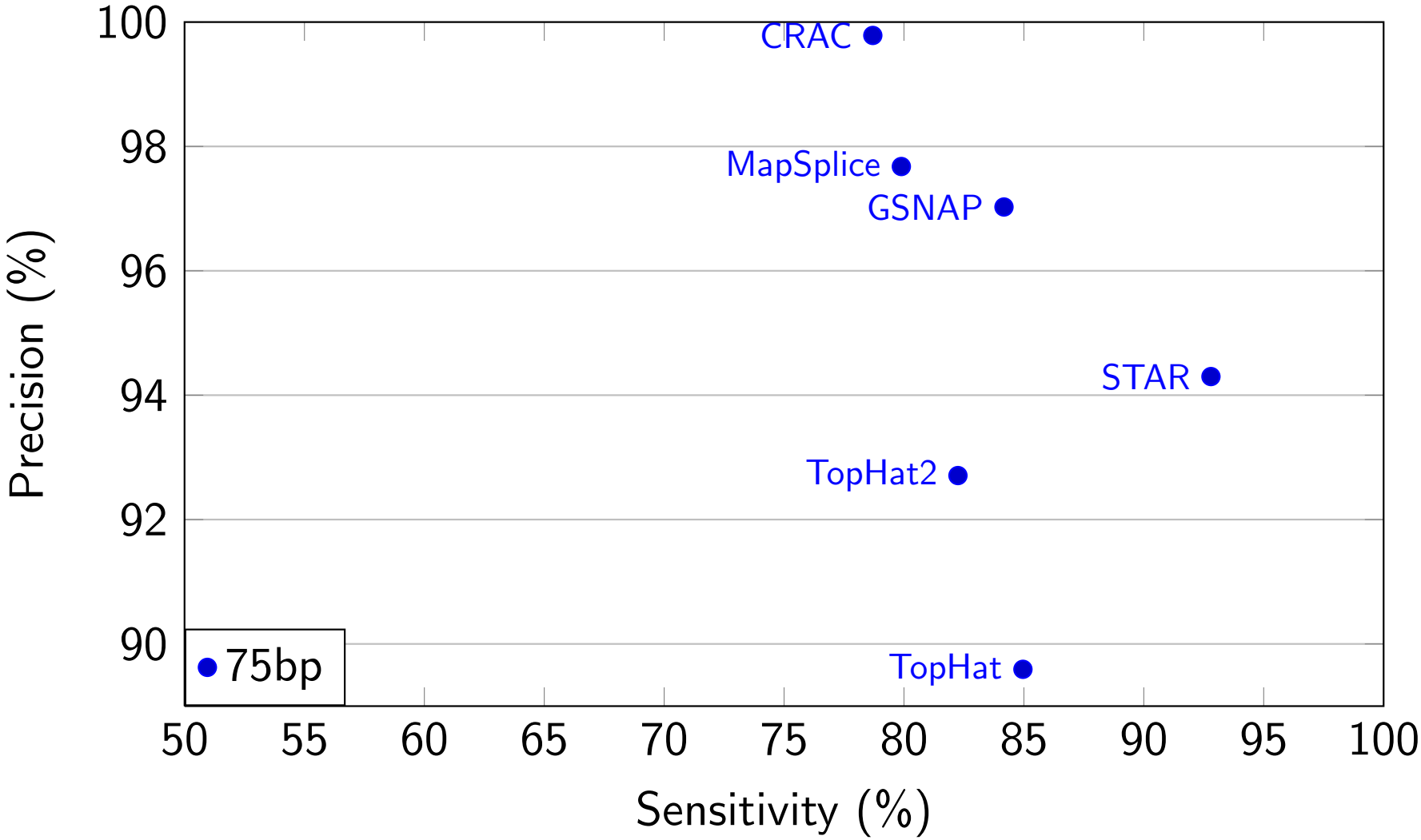
Alignment (single occurrences)



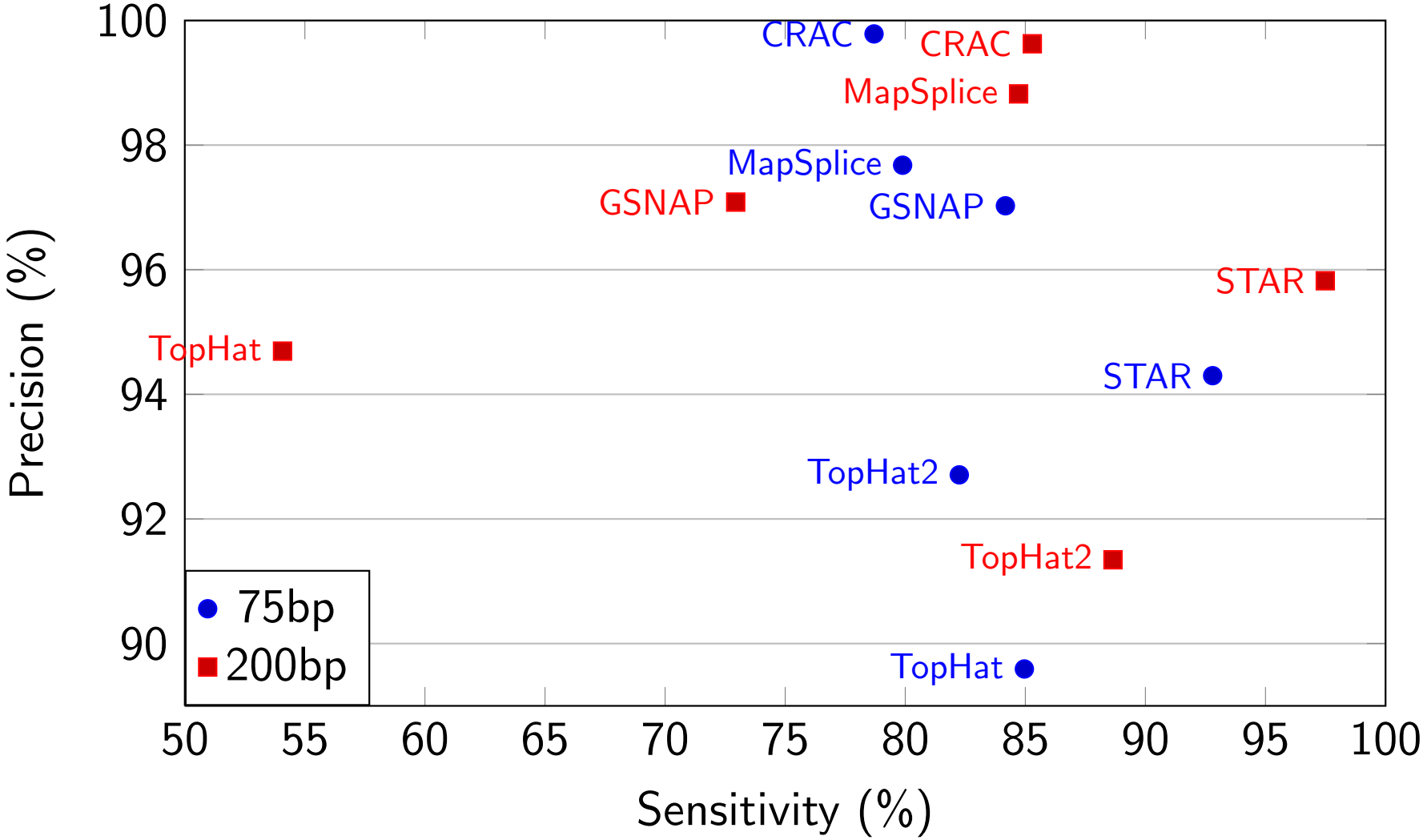
Alignment (single occurrences)



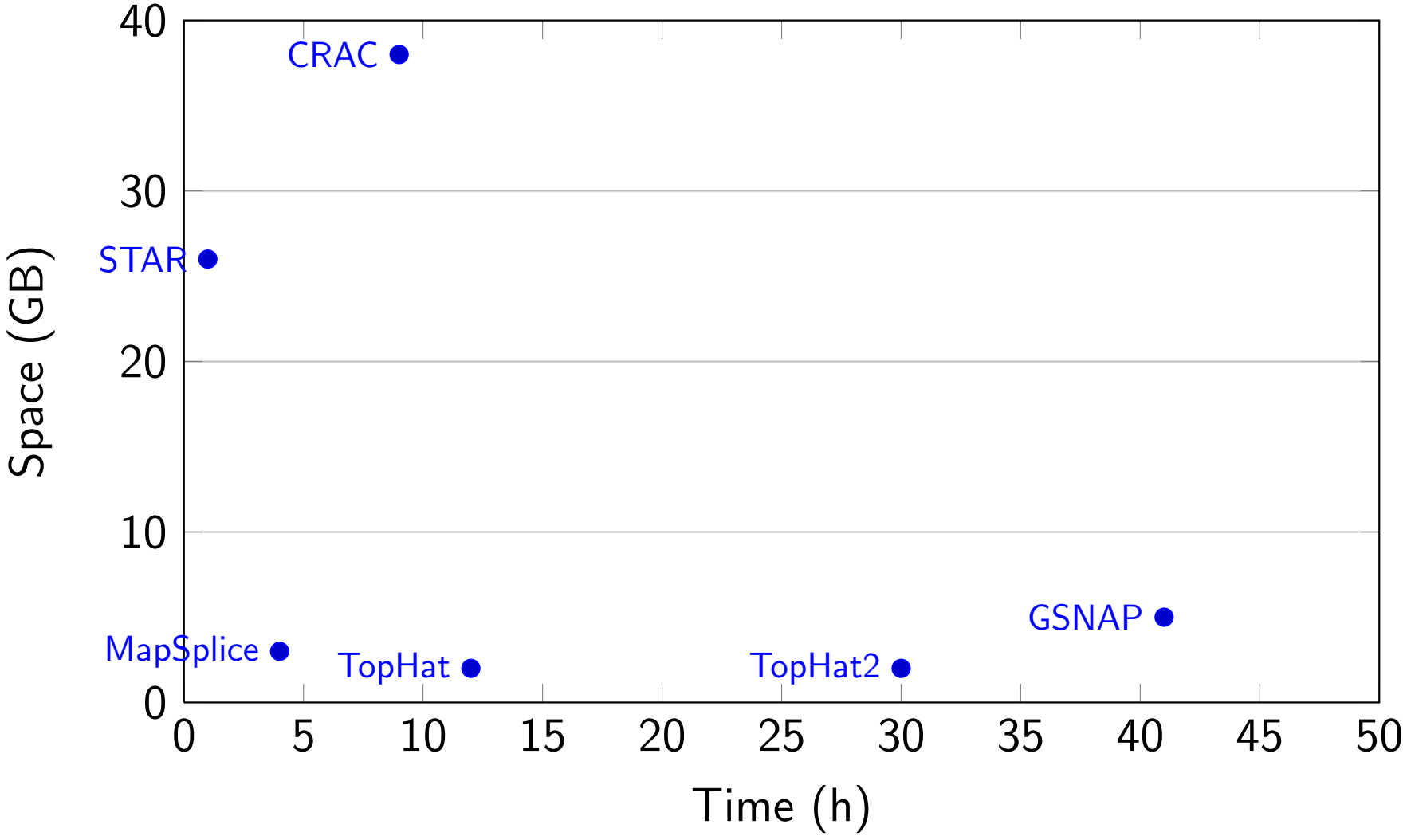
Exon-exon junction predictions



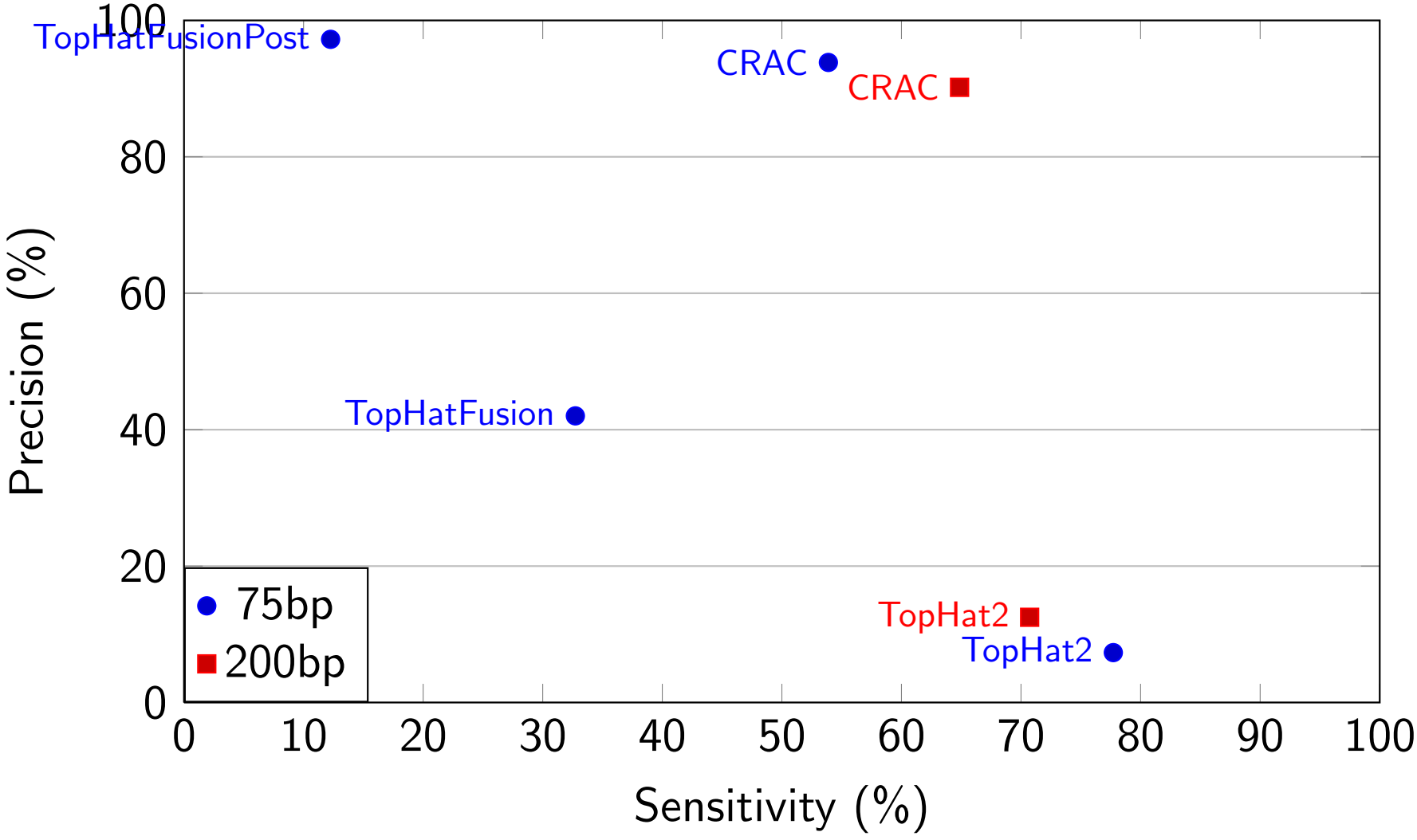
Exon-exon junction predictions



How many time? How many space?



Fusion predictions



Junction prediction in real data

ERR030856

RNA-Seq, Hi-Seq 2000
100bp stranded
76 M reads

Junction prediction in real data

ERR030856

RNA-Seq, Hi-Seq 2000
100bp stranded
76 M reads

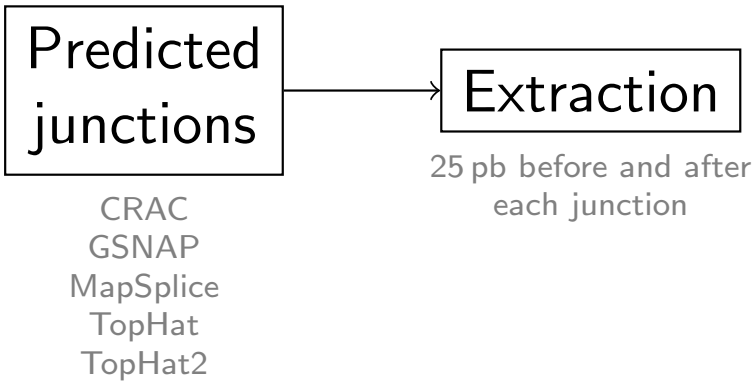
Predicted junctions

CRAC
GSNAP
MapSplice
TopHat
TopHat2

Junction prediction in real data

ERR030856

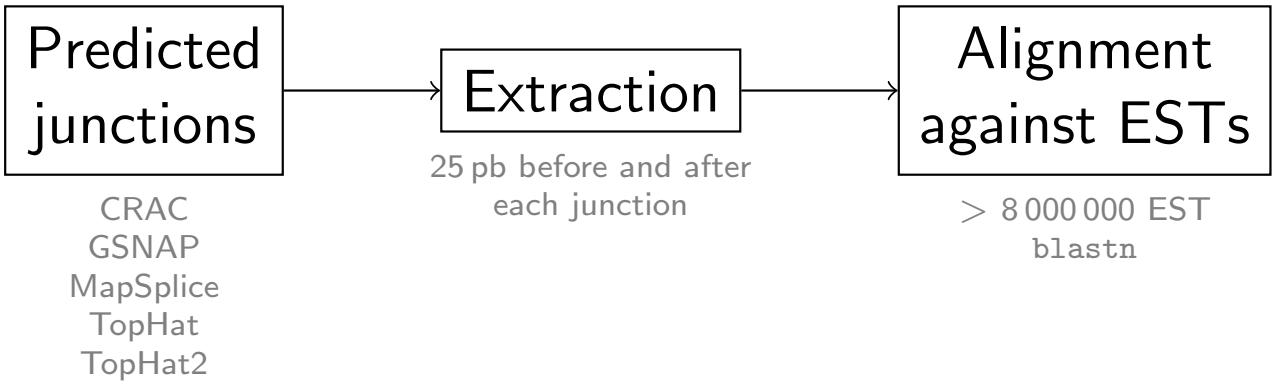
RNA-Seq, Hi-Seq 2000
100bp stranded
76 M reads



Junction prediction in real data

ERR030856

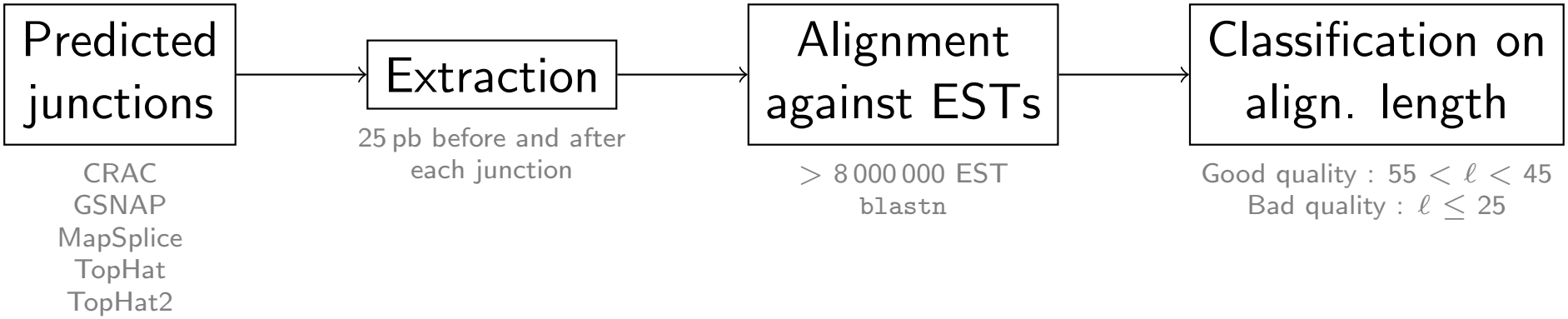
RNA-Seq, Hi-Seq 2000
100bp stranded
76 M reads



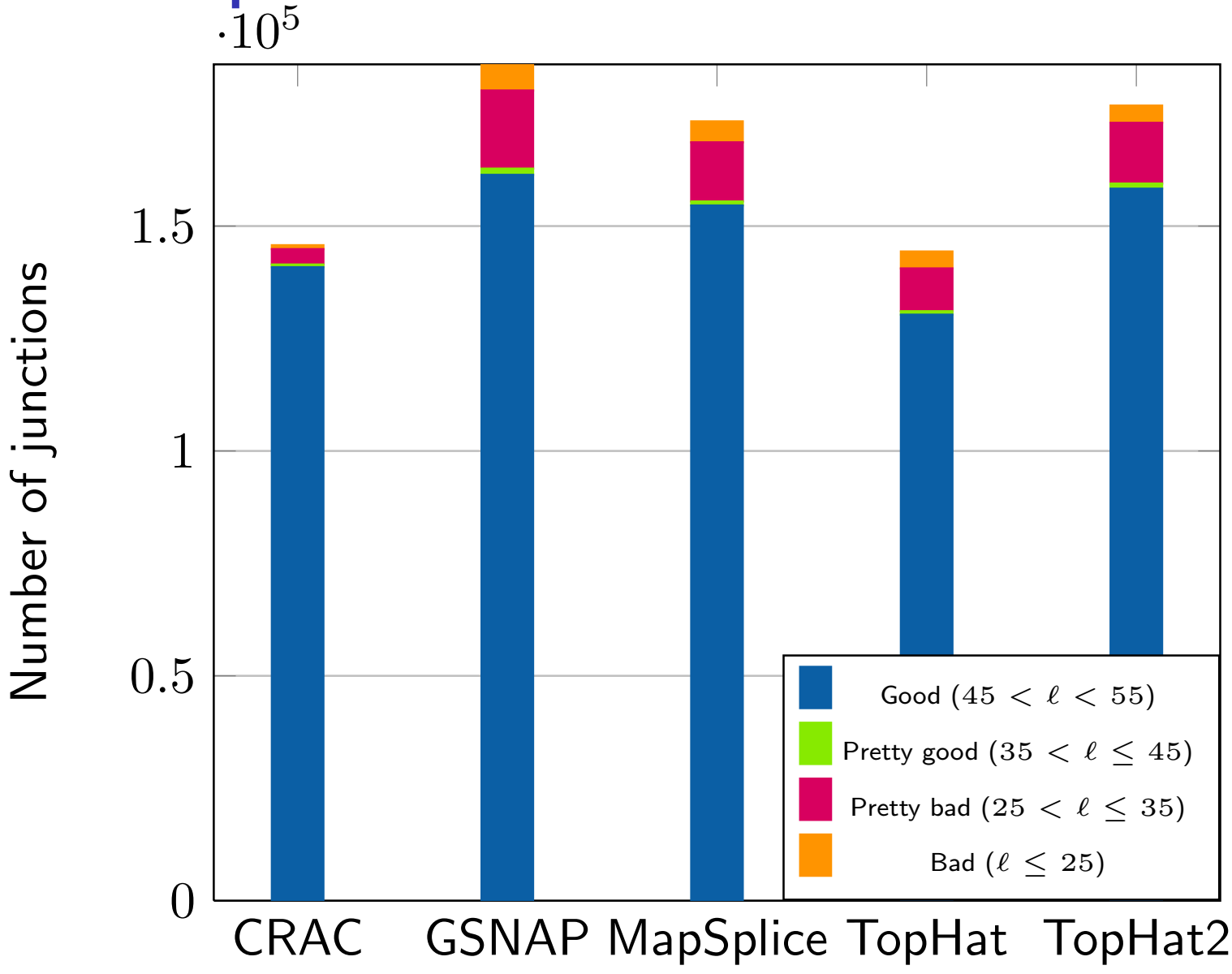
Junction prediction in real data

ERR030856

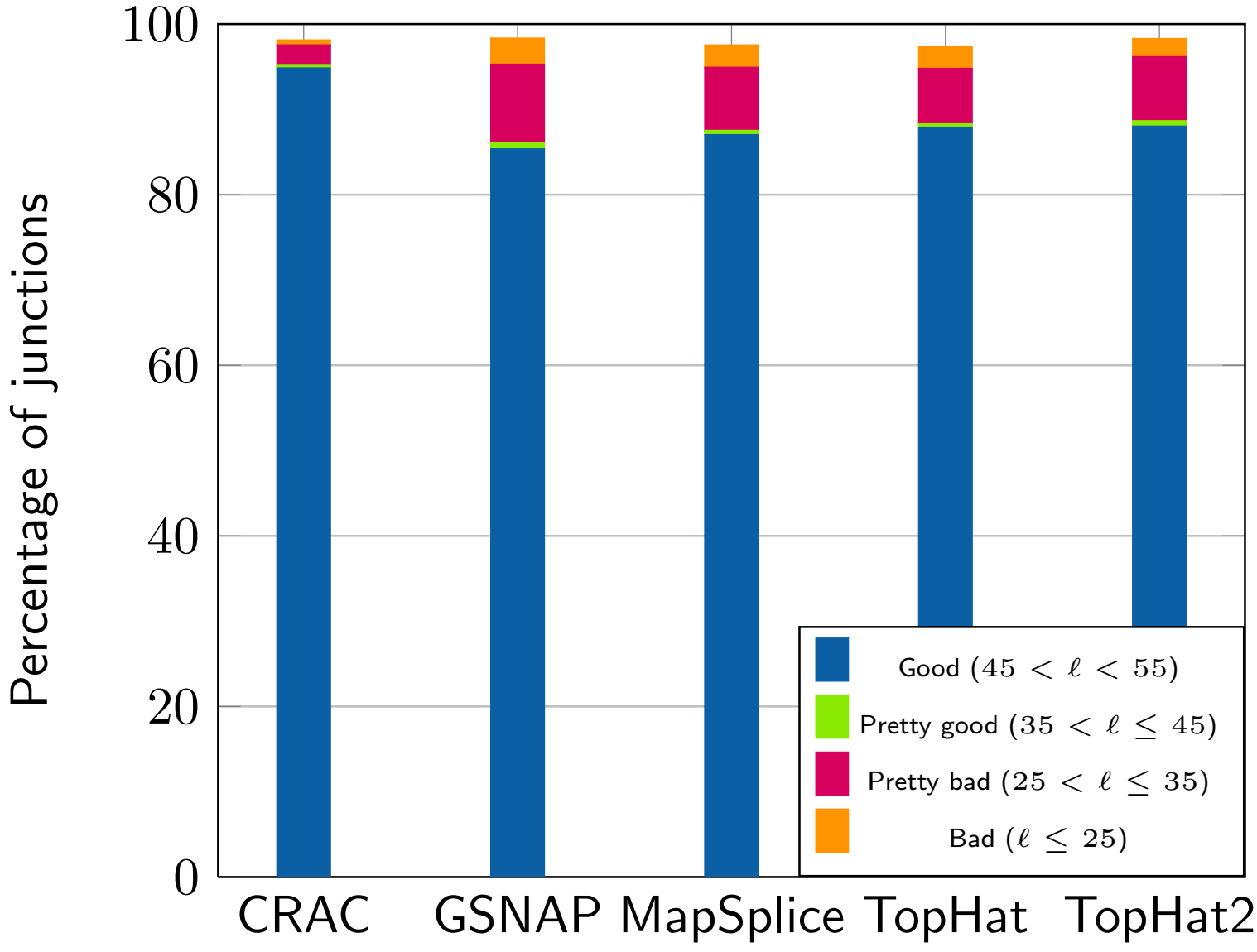
RNA-Seq, Hi-Seq 2000
100bp stranded
76 M reads



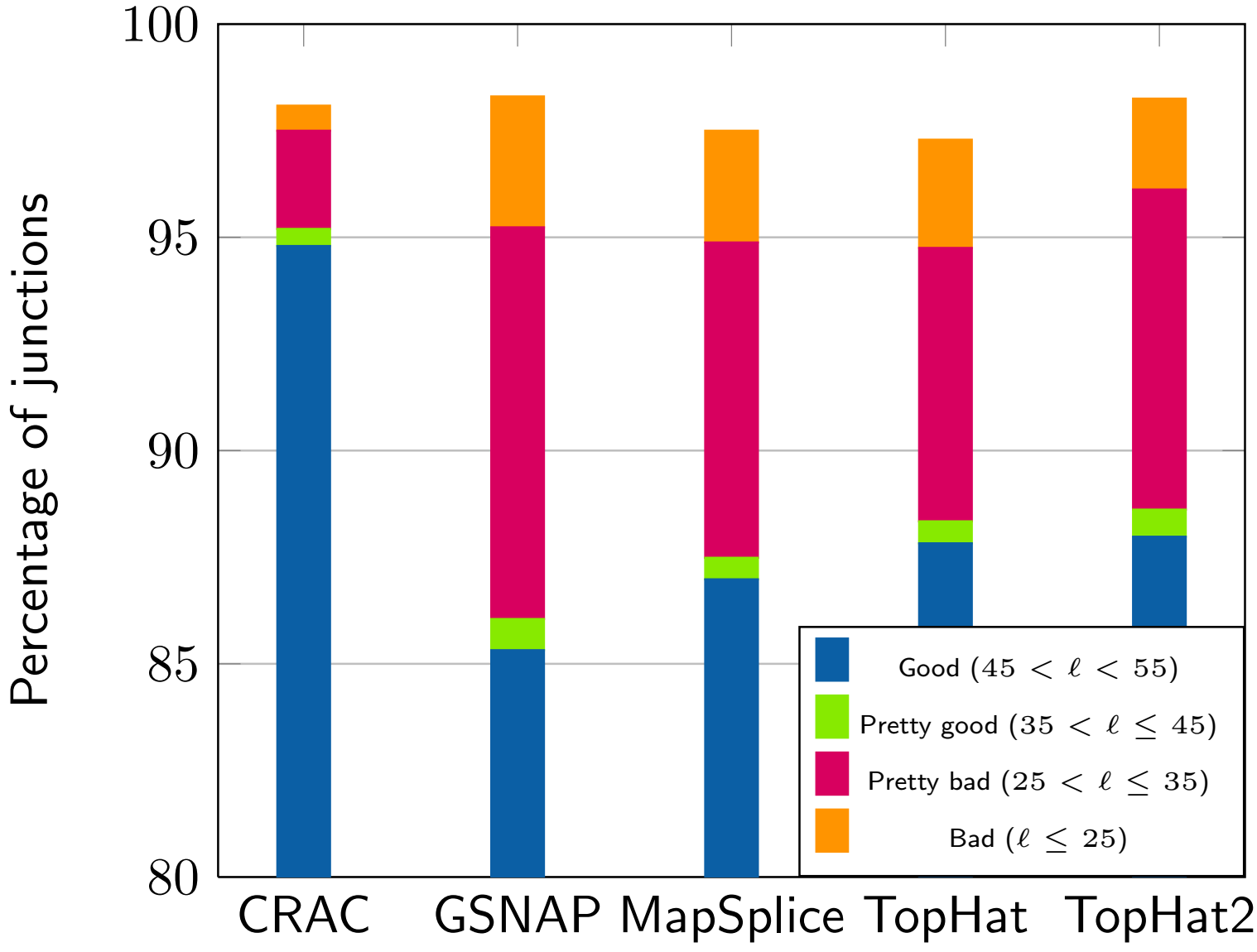
Junction prediction in real data



Junction prediction in real data



Junction prediction in real data

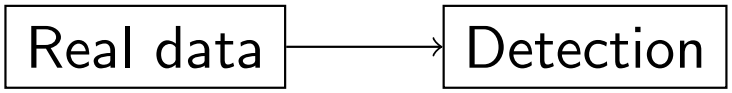


Fusions in real data

Real data

Breast cancer lineage
27 known fusions
50bp reads
Edgren *et al*, Genome Biology, 2011

Fusions in real data



Breast cancer lineage
27 known fusions
50bp reads
Edgren *et al*, Genome Biology, 2011

TopHatFusion
CRAC

Fusions in real data

Breast cancer lineage
27 known fusions
50bp reads
Edgren *et al*, Genome Biology, 2011

TopHatFusion
CRAC

Fusions in real data

Breast cancer lineage
27 known fusions
50bp reads
Edgren *et al*, Genome Biology, 2011

TopHatFusion
CRAC

CRAC

455

Fusions in real data

Breast cancer lineage
27 known fusions
50bp reads
Edgren *et al*, Genome Biology, 2011

TopHatFusion
CRAC

CRAC	**TopHatFusion**
455	193 163

Fusions in real data

Breast cancer lineage
27 known fusions
50bp reads
Edgren *et al*, Genome Biology, 2011

TopHatFusion
CRAC

CRAC	**TopHatFusion**
455	193 163
Validated: 20	Validated: 21

rac.

RNA-SEQ Mapping & Analysis

Jérôme Audoux

rac.

RNA-SEQ Mapping & Analysis

Jérôme Audoux

Re-sequencing

rac.

RNA-SEQ Mapping & Analysis

Jérôme Audoux

Re-sequencing

Unique location

rac.

RNA-SEQ Mapping & Analysis

Jérôme Audoux

Re-sequencing

Unique location

Long reads

rac.

RNA-SEQ Mapping & Analysis

Jérôme Audoux

http://crac.gforge.inria.fr

Re-sequencing

Unique location

Long reads