

BIOINFORMATICS
INSTITUTE

Primer Trimming for Targeted Resequencing

Timofey Prodanov

Maxim Ivanov, Atlas

Vladislav Mileiko, Atlas

Problem



	Allele 1	●	
	Allele 2		



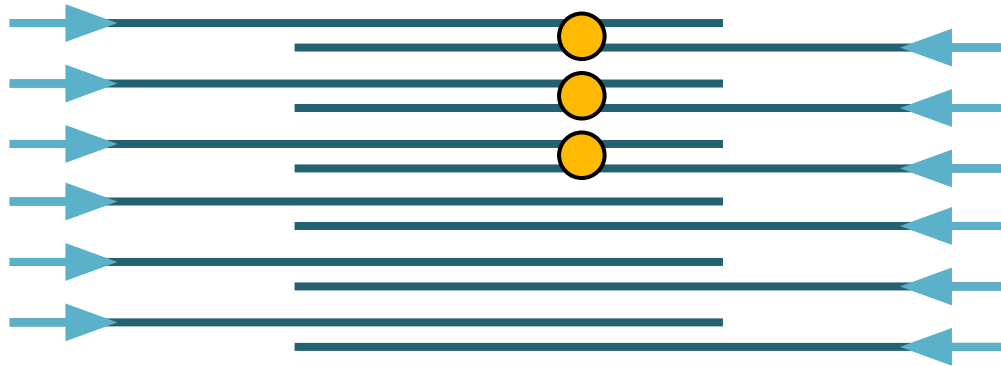
Problem



Allele 1



Allele 2



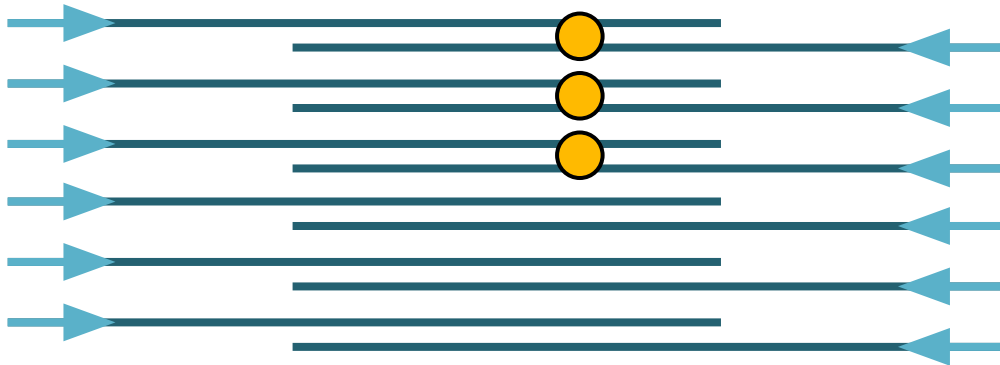
Problem



Allele 1



Allele 2



Amplicon 1

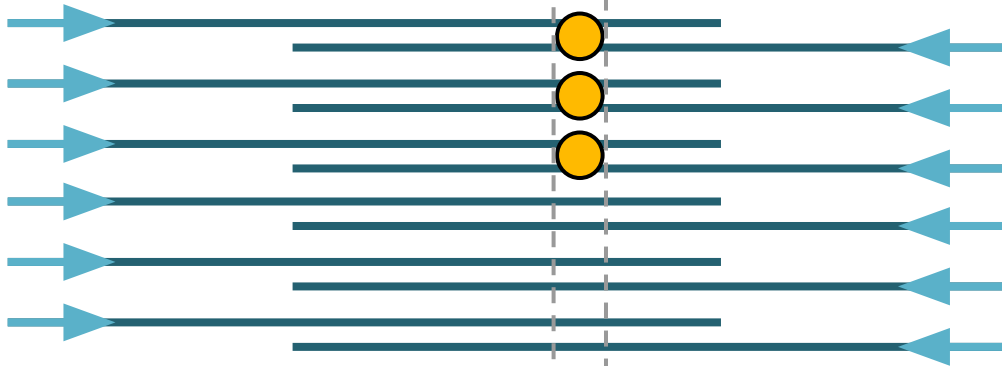


Amplicon 2

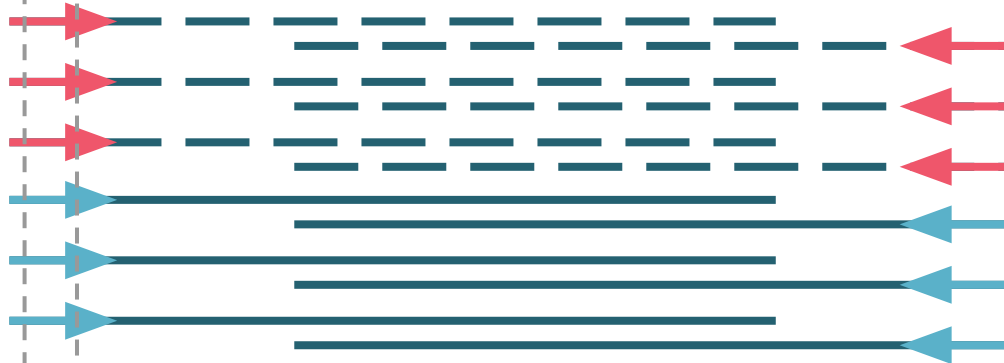
Problem



Variant A



Amplicon 1

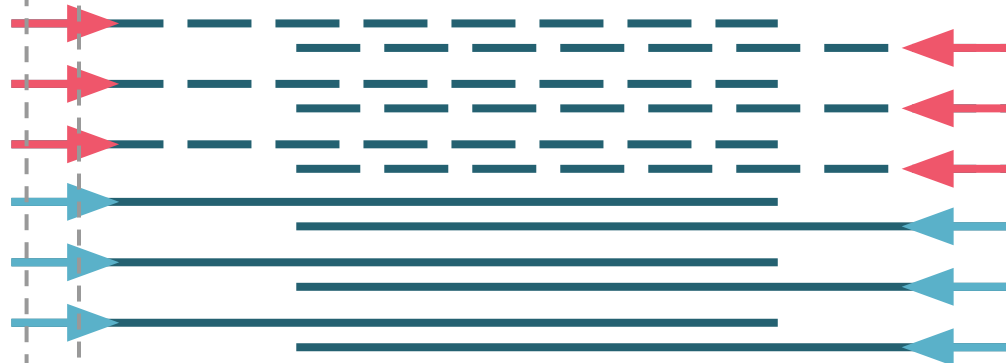
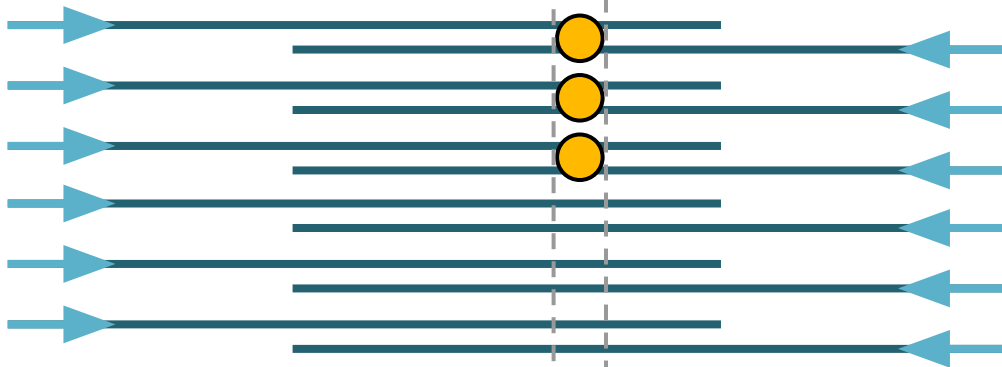


Amplicon 2

Problem



Variant A



Allele specific drop-out:
Variant A: $6 / 15 = 40\%$

Amplicon 2

Problem

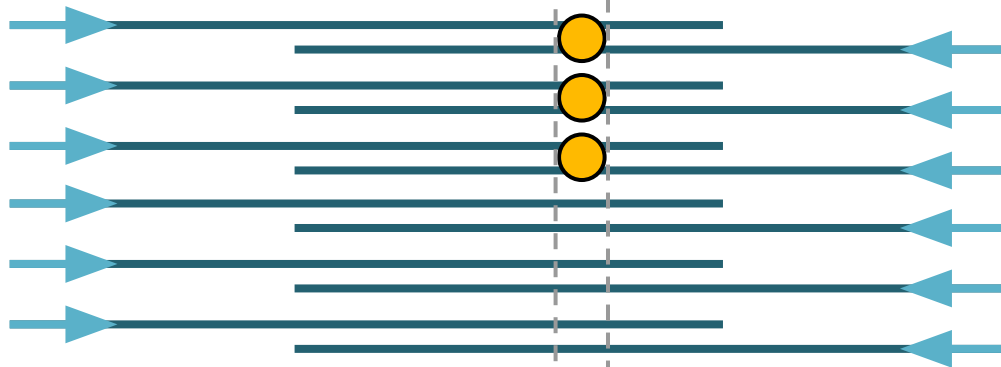


Variant A

Allele 1



Allele 2



Amplicon 1



Amplicon 2

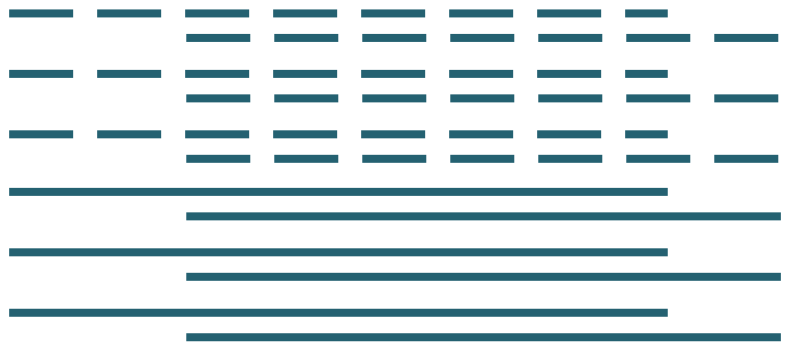
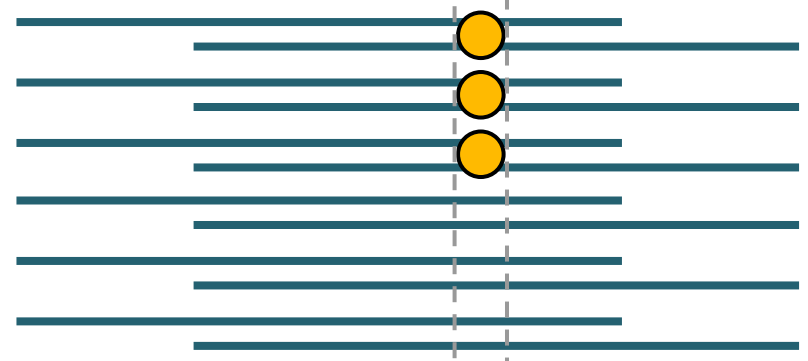
Without drop-out:

Variant A: $6 / 18 = 33\%$

Solution



Variant A

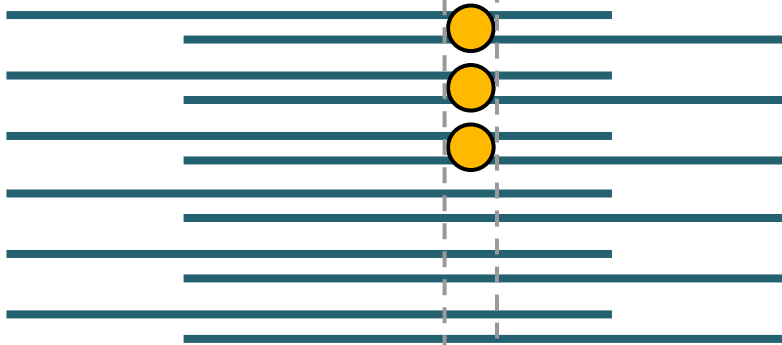


Amplicon 2

Solution



Variant A



Allele specific drop-out:
Variant A: $6 / 12 = 50\%$

Amplicon 2

Solution

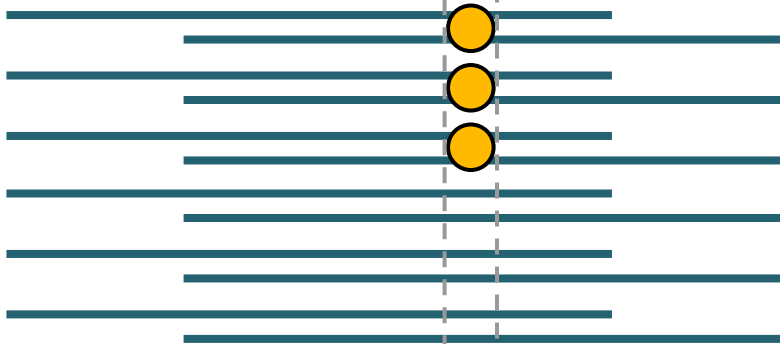


Variant A

Allele 1



Allele 2



Amplicon 1

Without drop-out:

Variant A: $6 / 12 = 50\%$



Amplicon 2

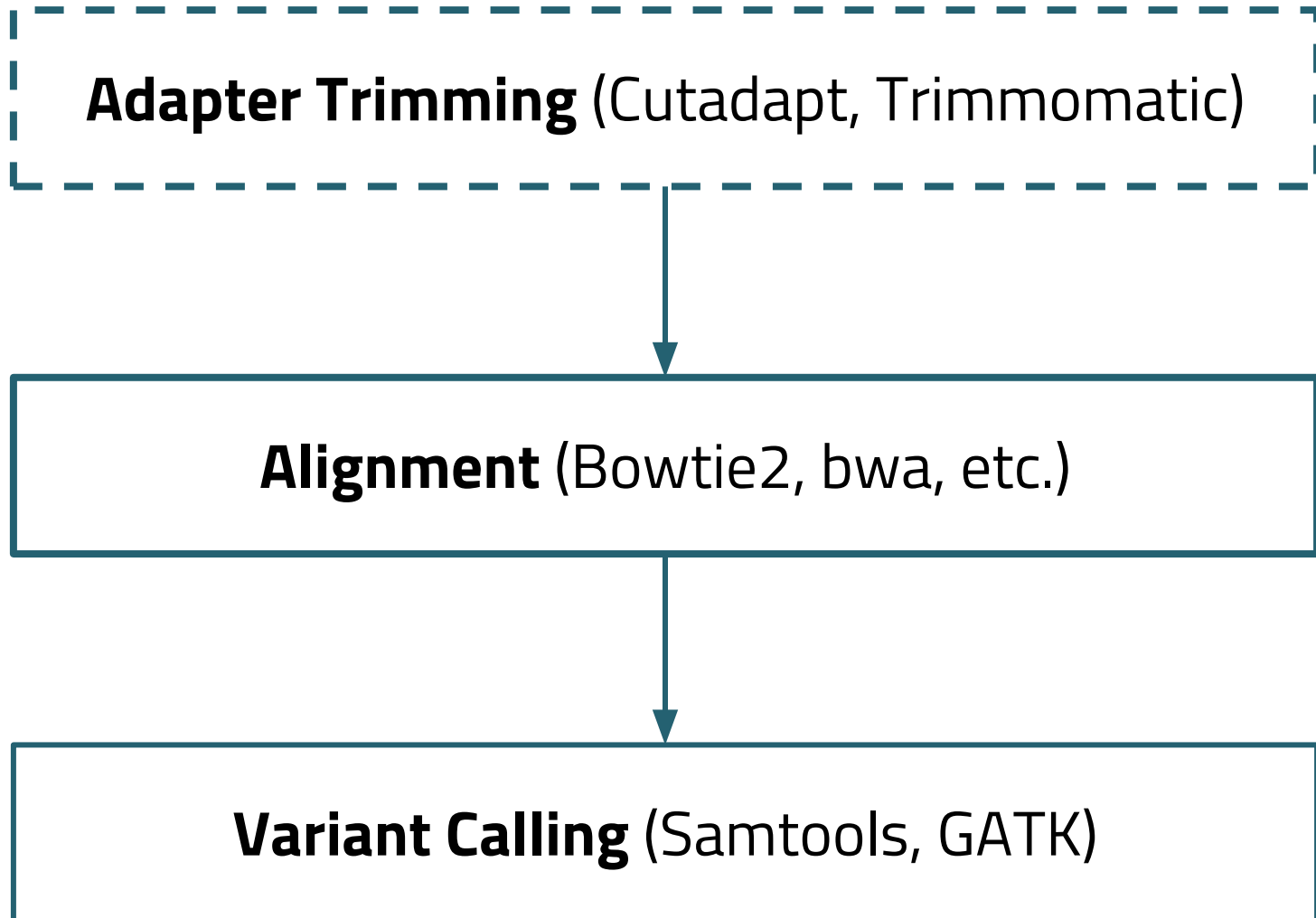
More about the problem

9 patients with causative mutation **E92K**
(4th exon of the CFTR)

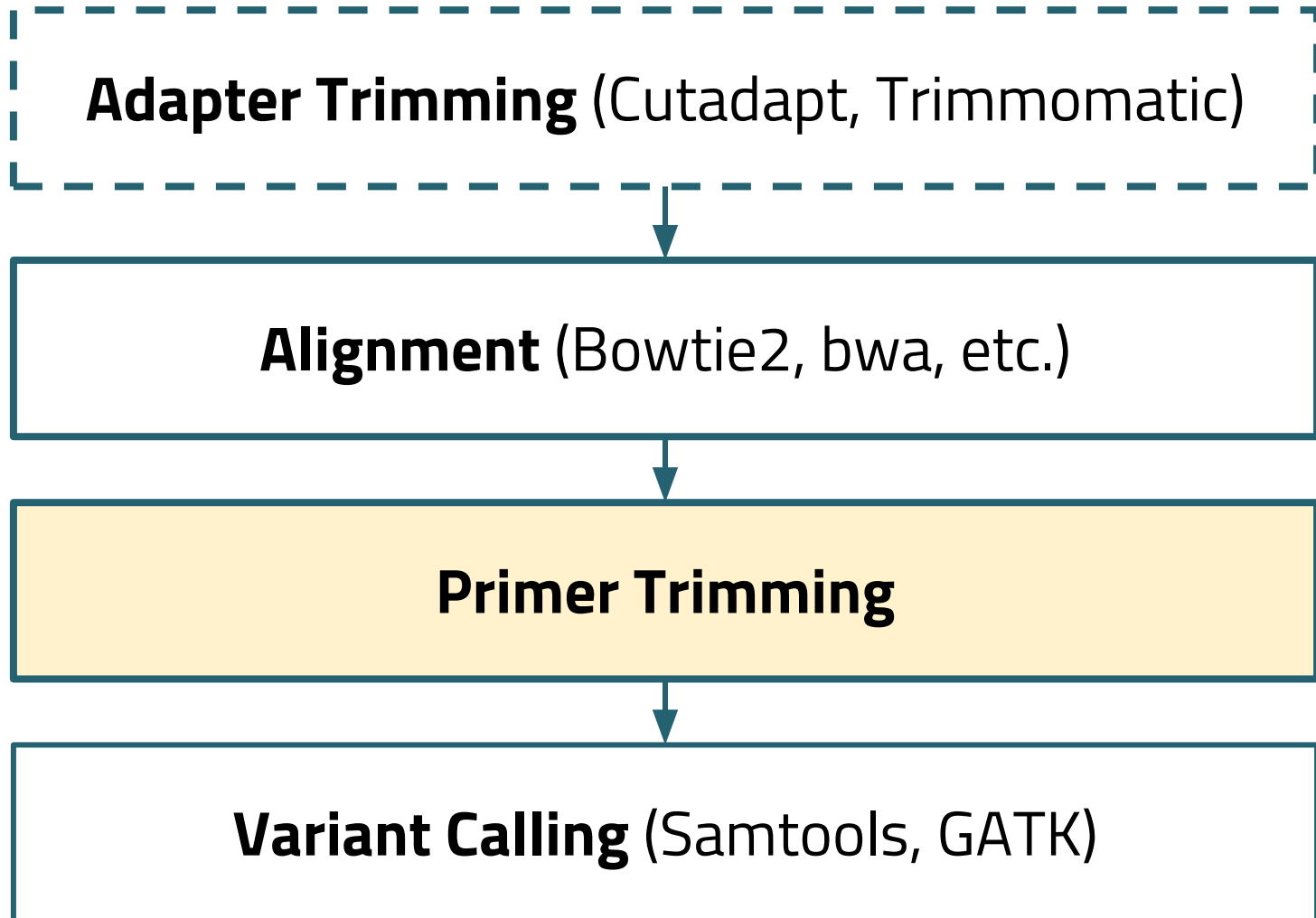
The mutation collocated with a primer hybridization site

Samtools and GATK **could not detect** this mutation

Suggested pipeline



Suggested pipeline



Primer trimming: Unpaired reads

Unpaired reads



Amplicons (.fasta)



Primer trimming: Unpaired reads

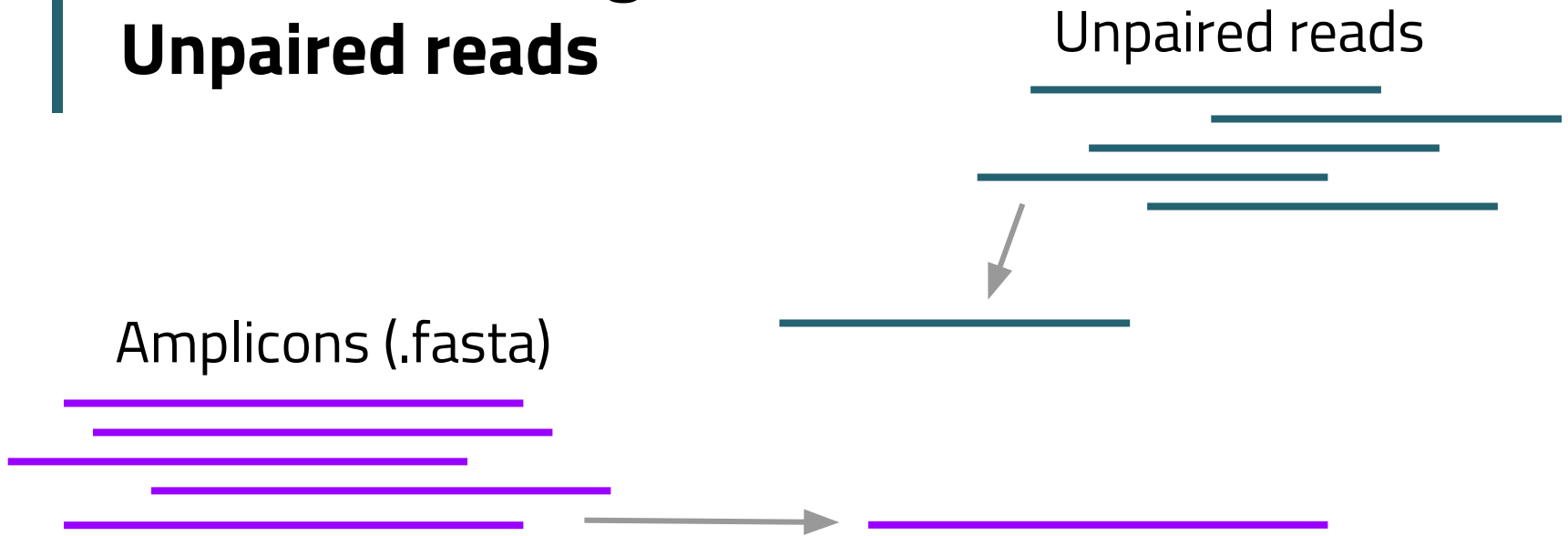
Amplicons (.fasta)



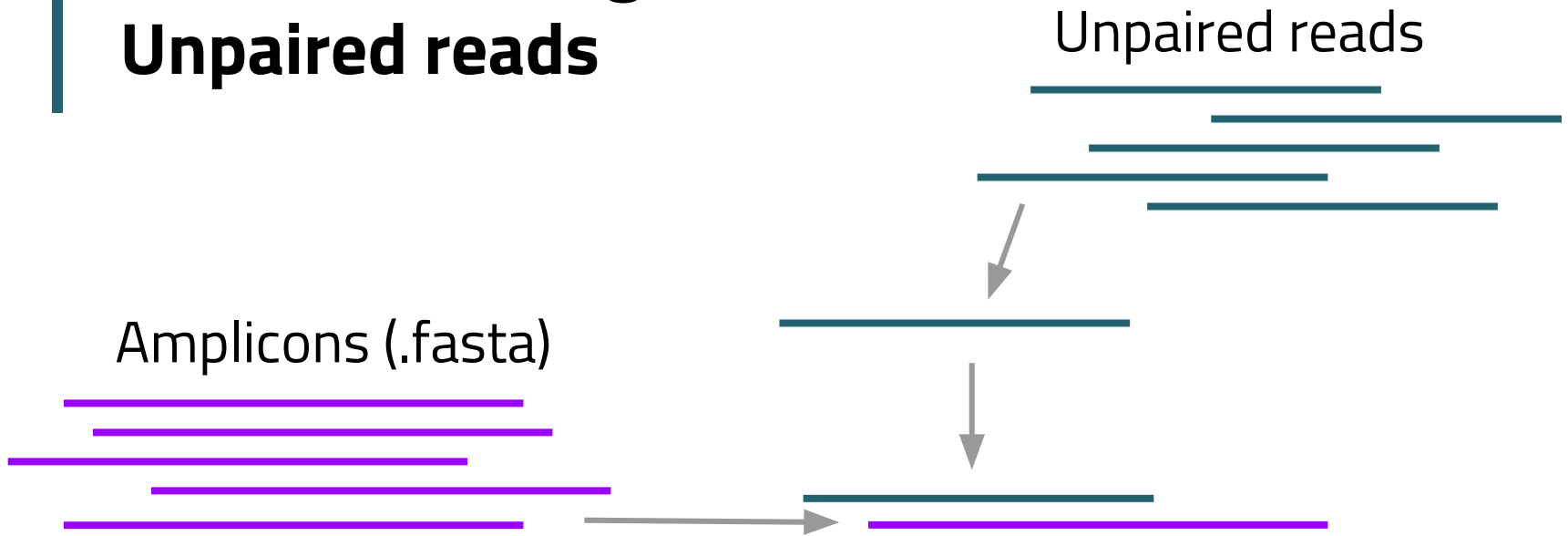
Unpaired reads



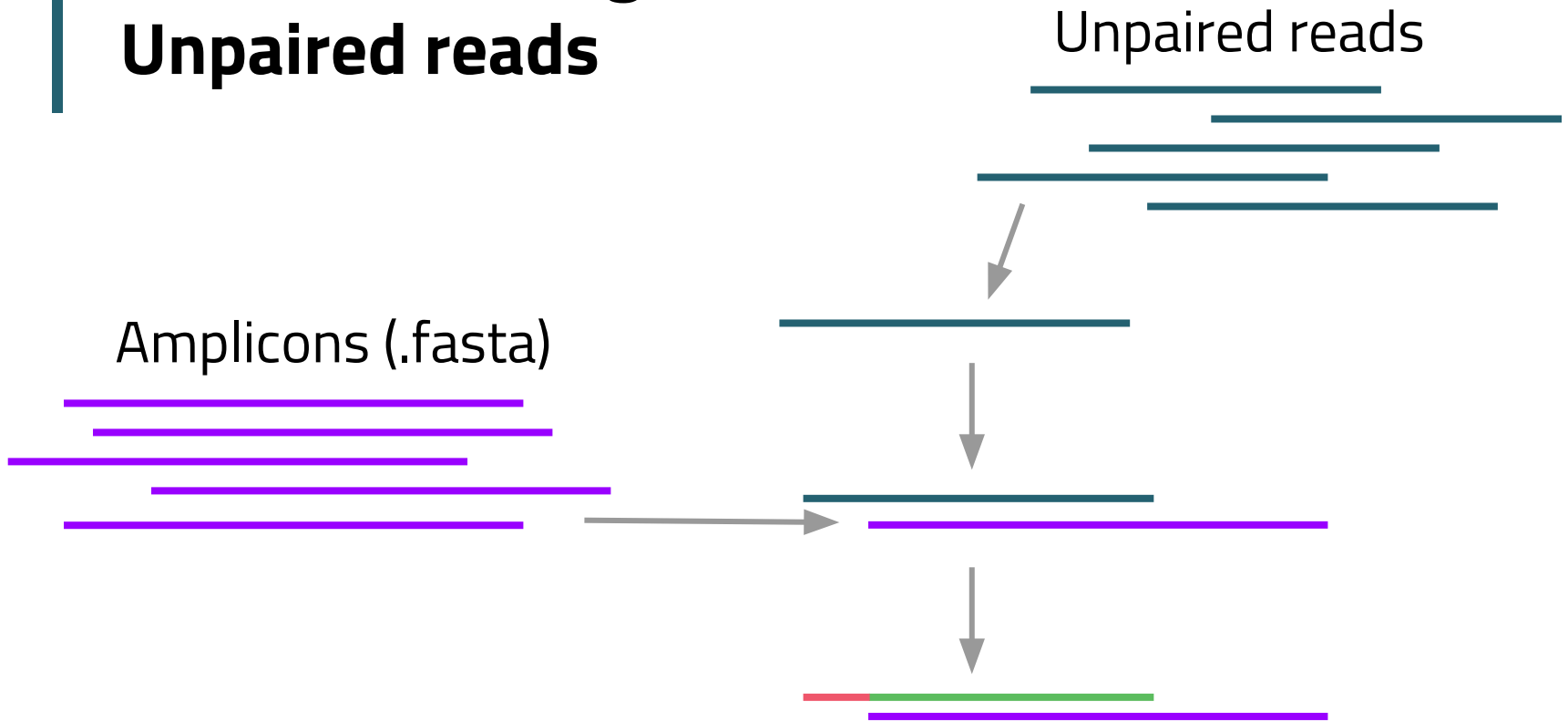
Primer trimming: Unpaired reads



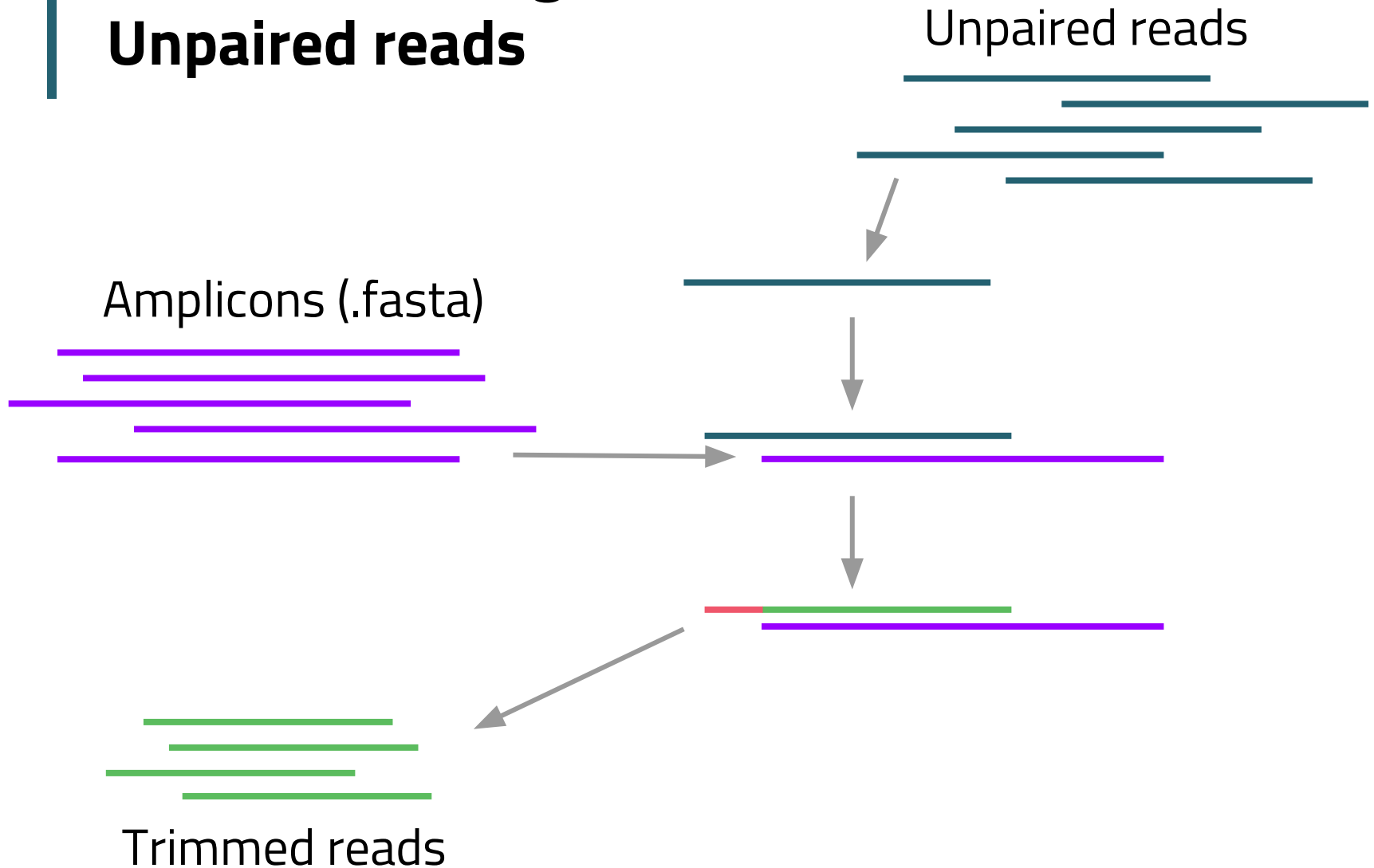
Primer trimming: Unpaired reads



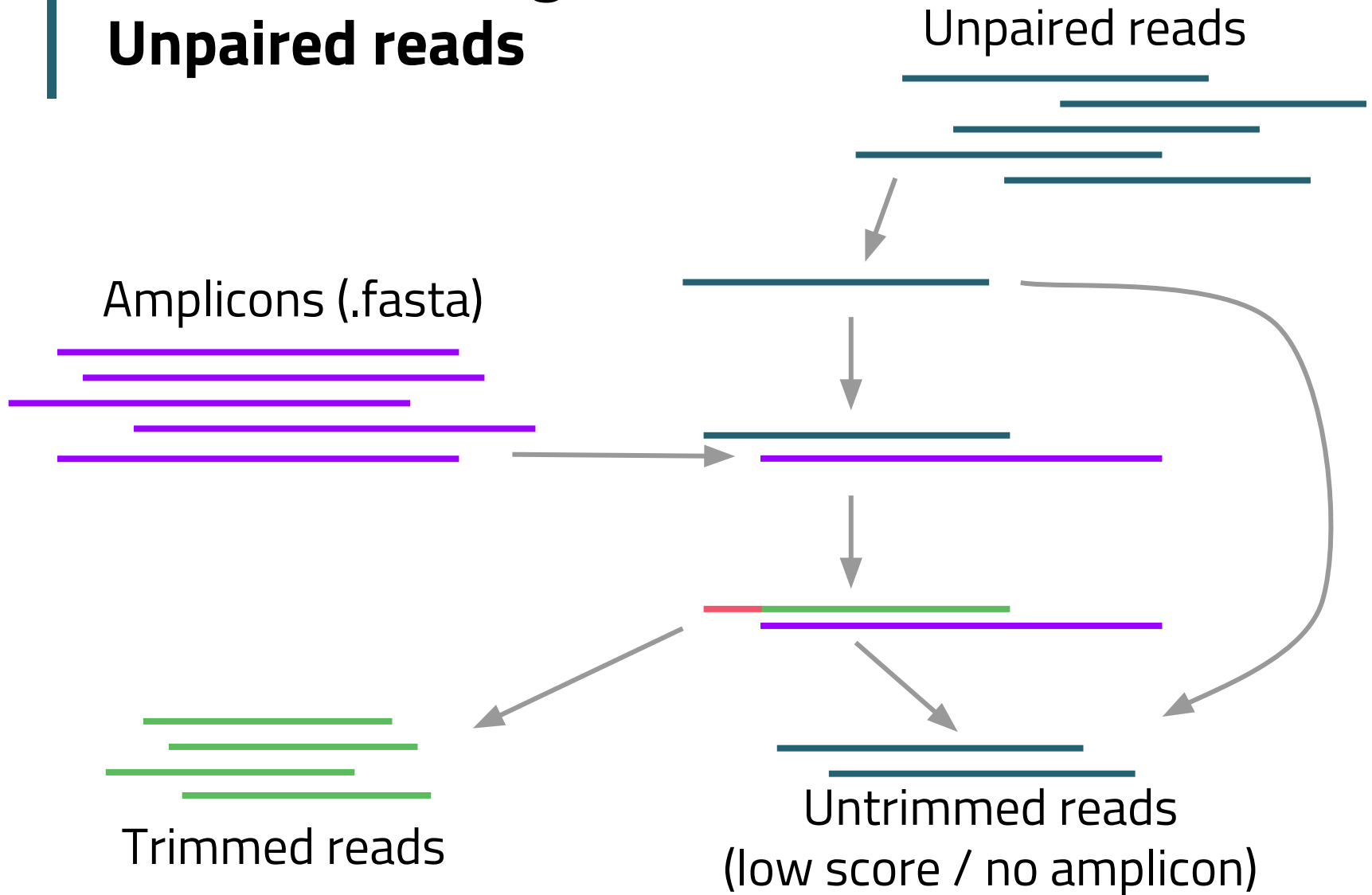
Primer trimming: Unpaired reads



Primer trimming: Unpaired reads



Primer trimming: Unpaired reads

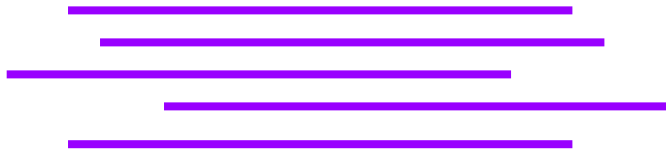


Primer trimming: Paired reads

Paired reads

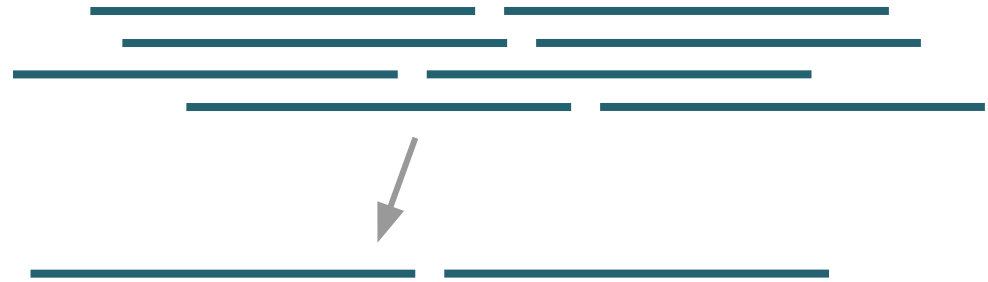


Amplicons (.fasta)



Primer trimming: Paired reads

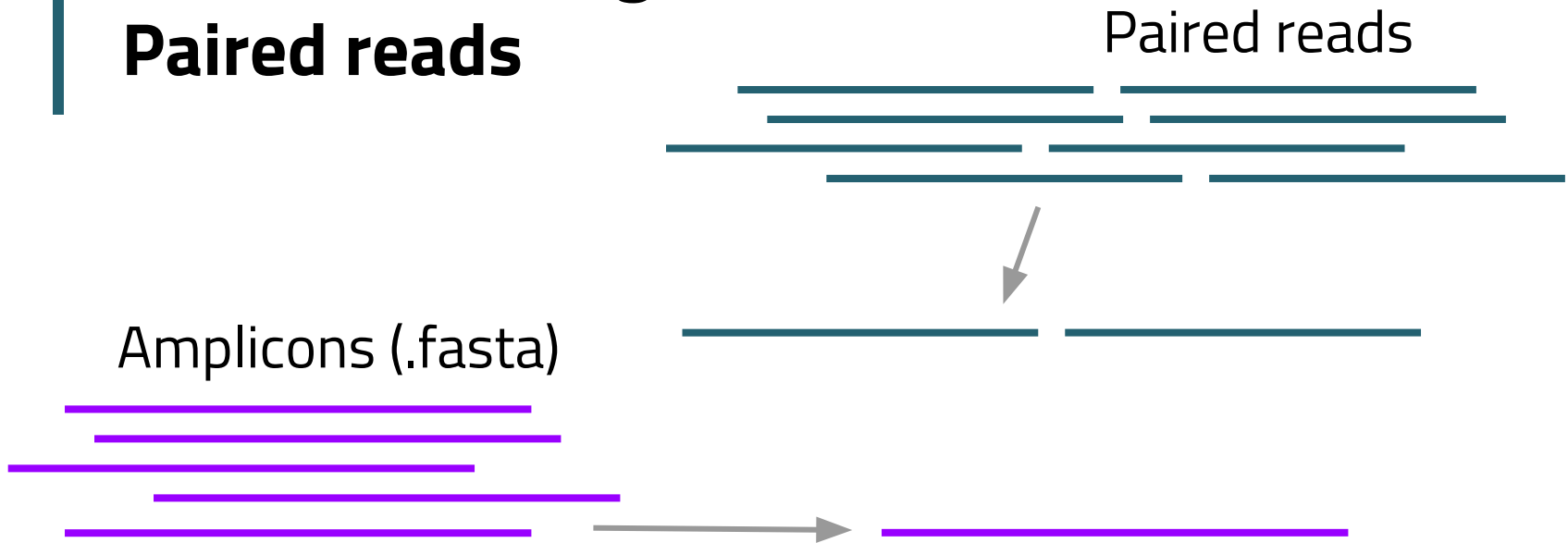
Paired reads



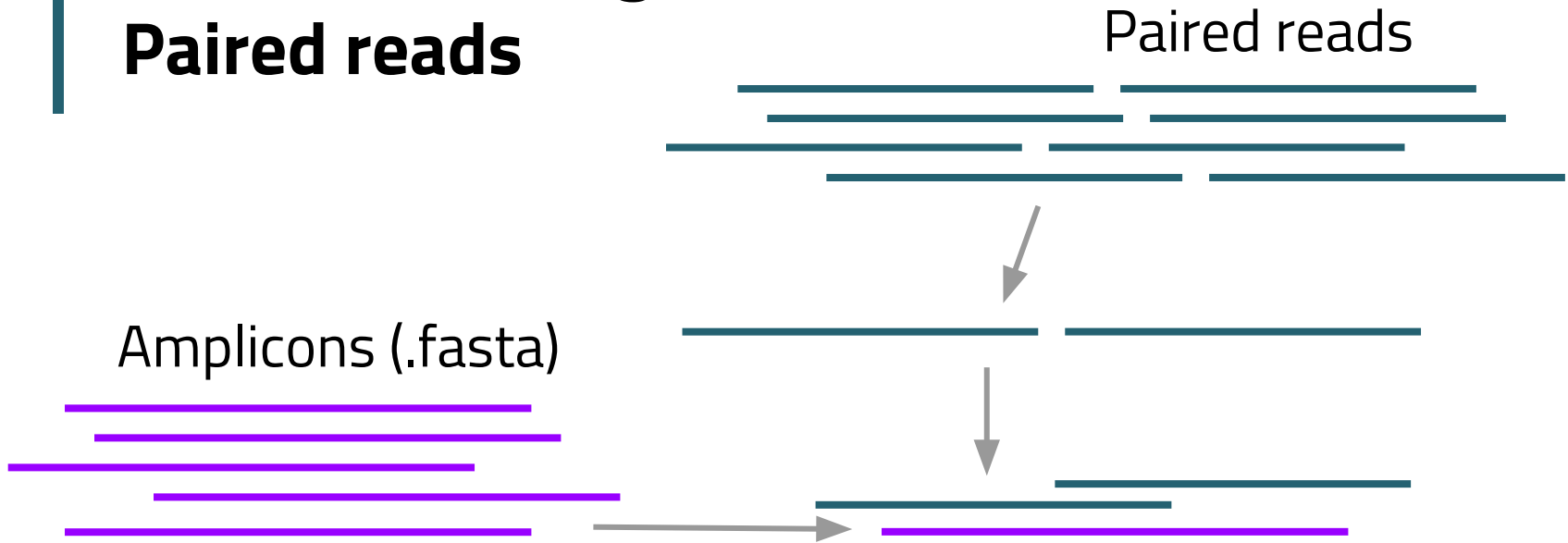
Amplicons (.fasta)



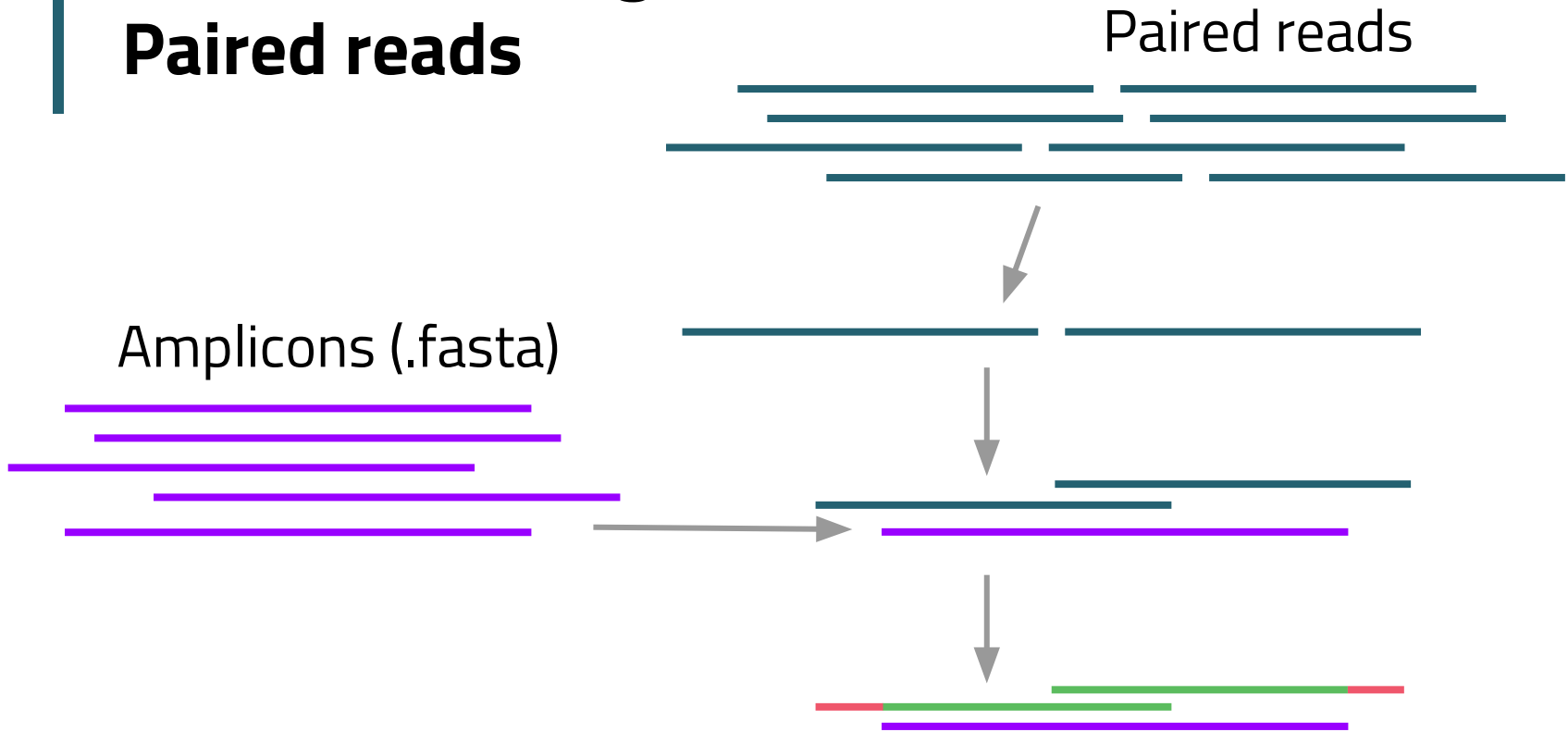
Primer trimming: Paired reads



Primer trimming: Paired reads



Primer trimming: Paired reads

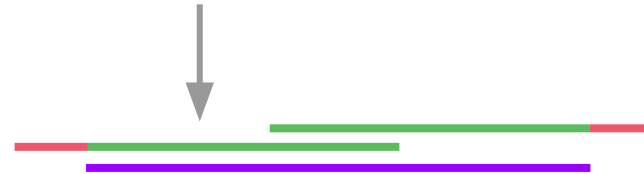
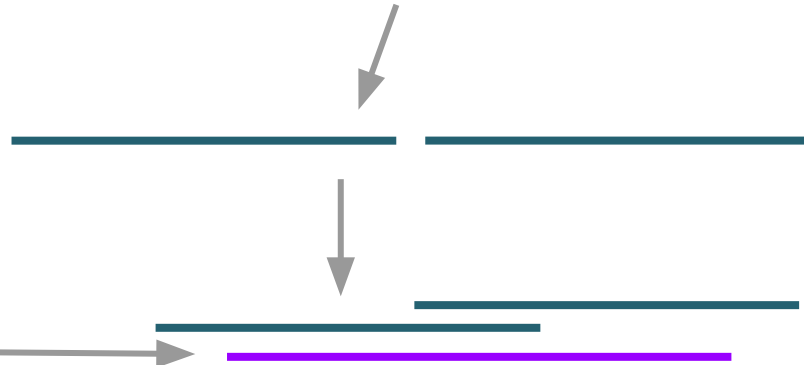


Primer trimming: Paired reads

Paired reads

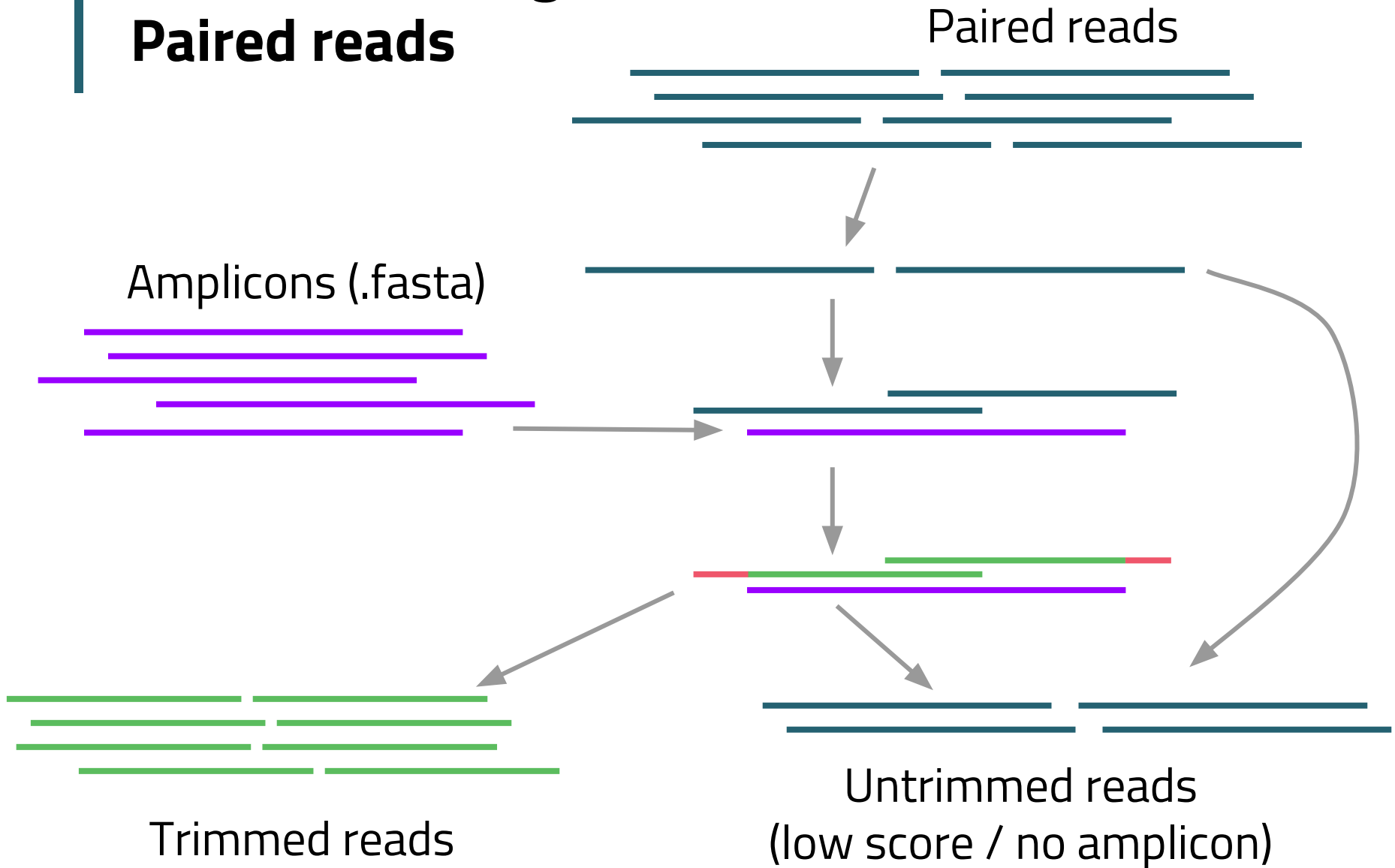


Amplicons (.fasta)



Trimmed reads

Primer trimming: Paired reads



Amplicon discovery

Small size of target reference \Rightarrow collisions are rare

Perform full k-mer decomposition of amplicons

At most 7 k-mer positions in every read

Amplicon discovery

k-mer positions in every read



Amplicon discovery

k-mer positions in every read



If no amplicons were found

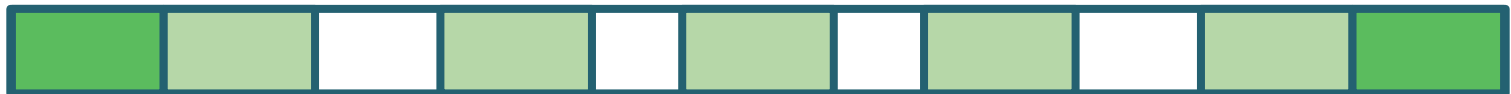


Amplicon discovery

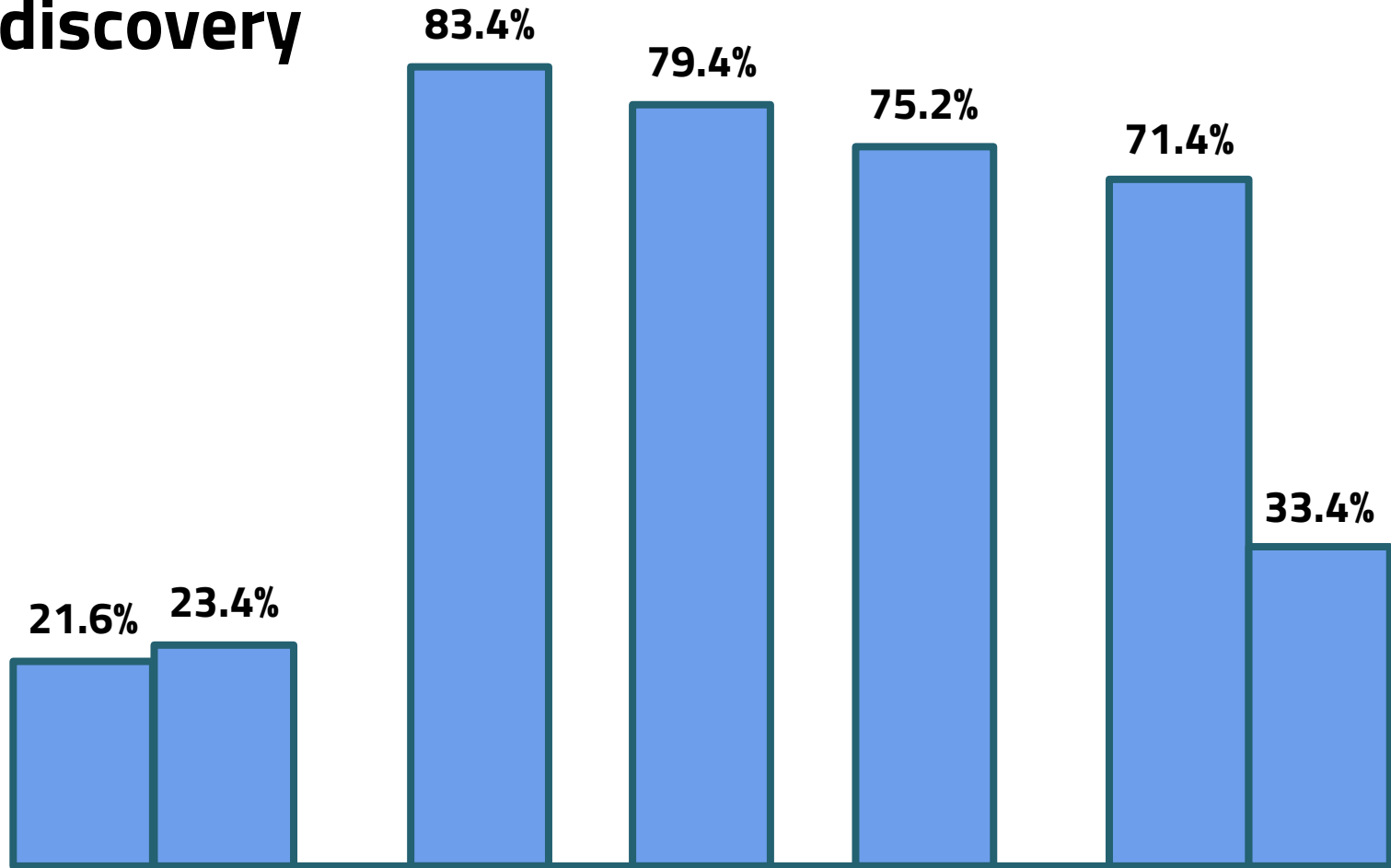
k-mer positions in every read



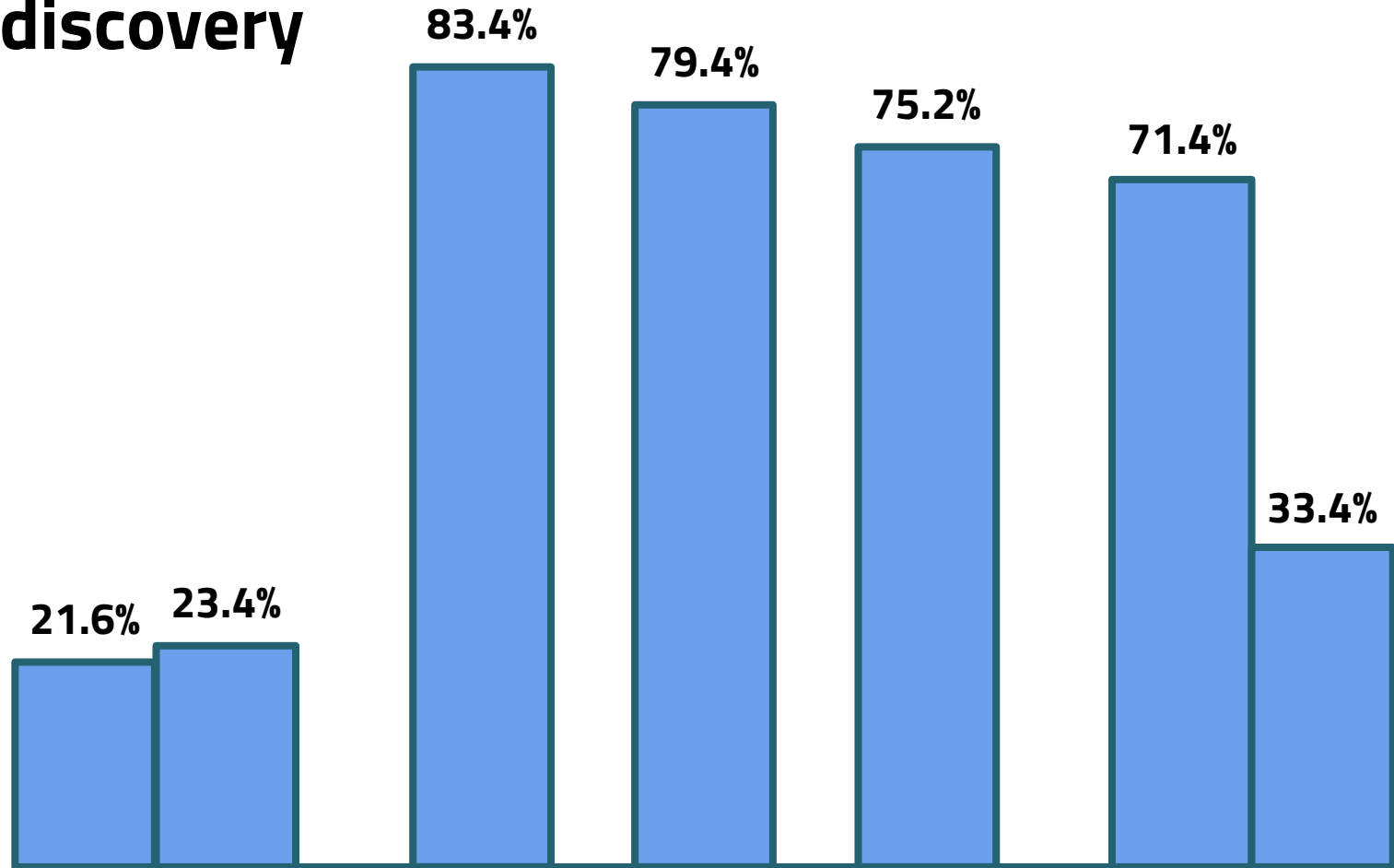
If no amplicons were found



Amplicon discovery



Amplicon discovery

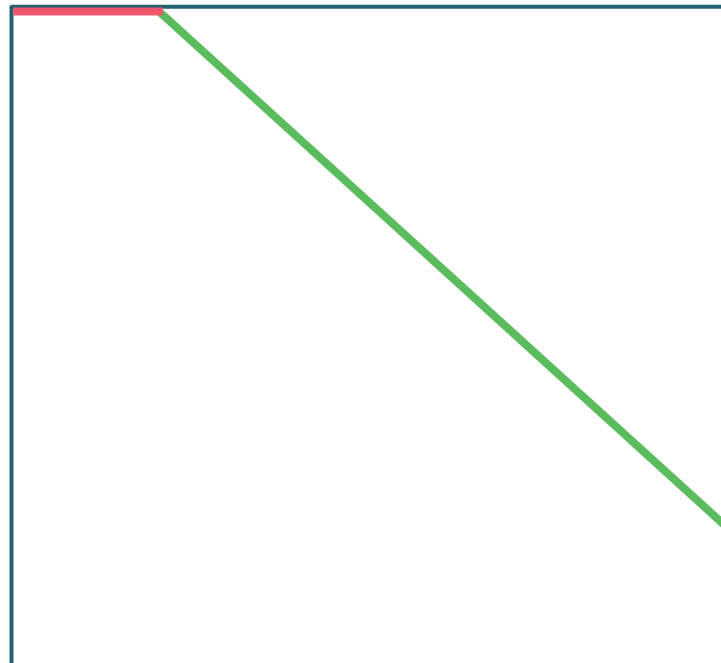


Unmatched: 7.3% (less than with bwa and Bowtie2)

Alignment: Single reads



Read

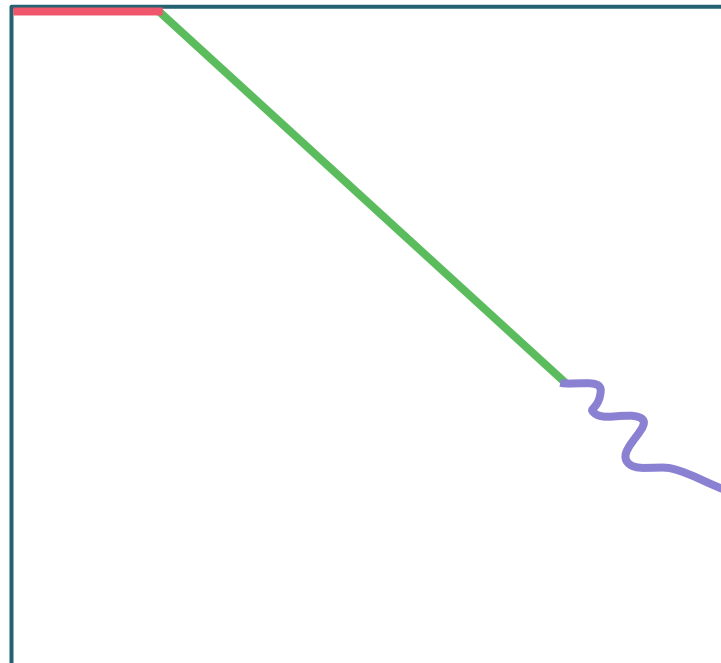


Just local
alignment

Alignment: Single reads



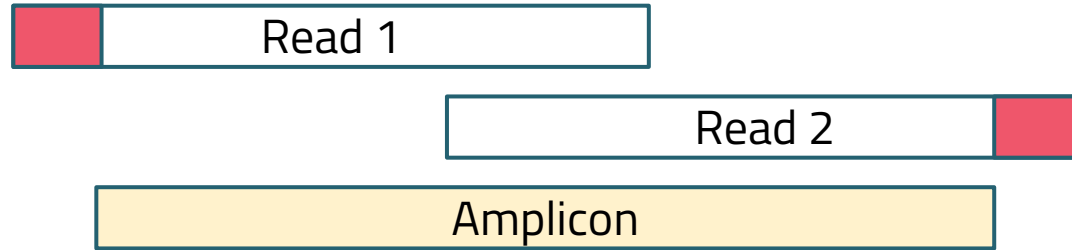
Read



Amplicon

Just local
alignment

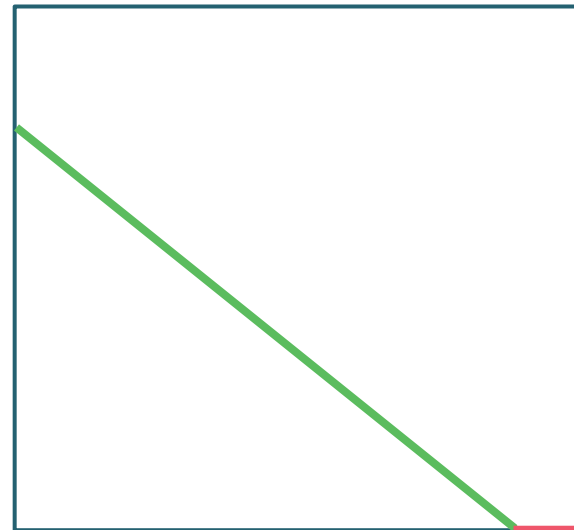
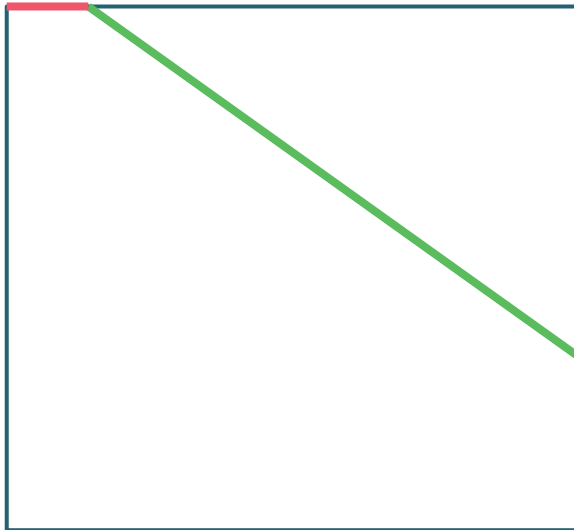
Alignment: Paired reads



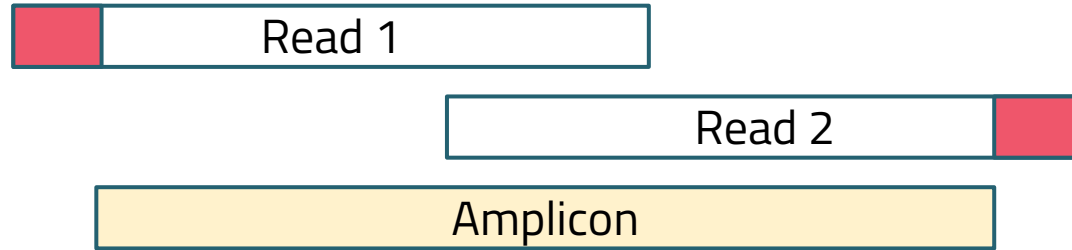
Read 1

Read 2

Amplicon



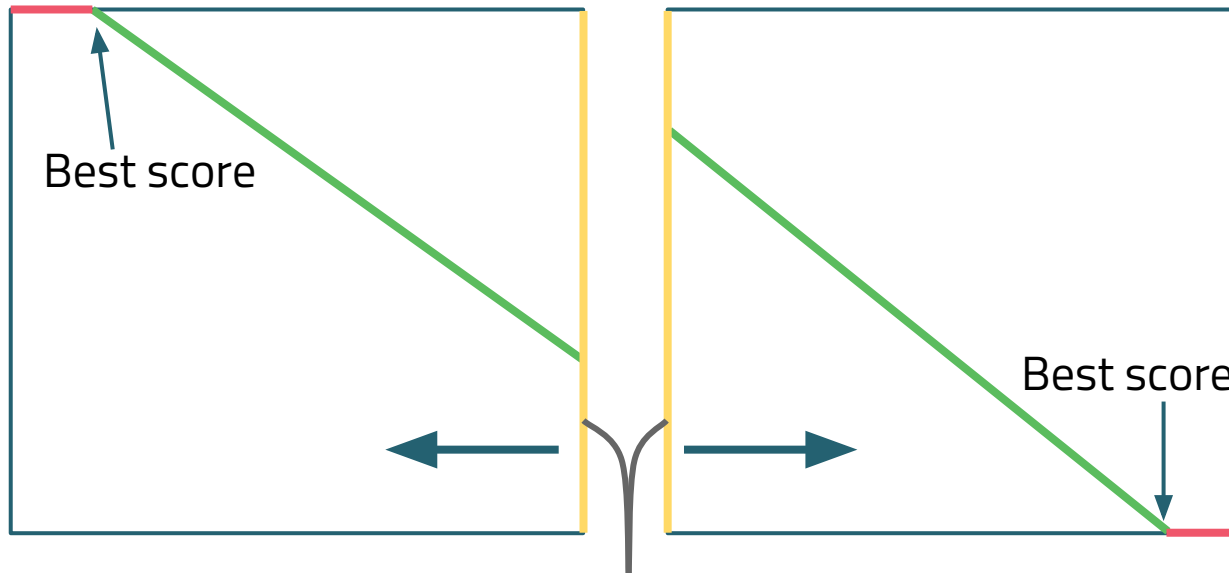
Alignment: Paired reads



Read 1

Read 2

Amplicon



Free rides area

Software product

- Named **Pretry**
- Easy to download and use
<https://github.com/tprodanov/Pretry>
- Methods description
- README
- Supported both .fastq and .sam formats
- .sam file remains consistent

PreTTY by Timofey Prodanov (timofey.prodanov@gmail.com)

Usage:

```
preTTY [options] -a <ampl.fa> -f [<in.fq>] [-o <out.fq>] [-e <err.fq>]
preTTY [options] -a <ampl.fa> -f <in1.fq> <in2.fq> -o <out1.fq> <out2.fq> [-e <err1.fq>]
preTTY [options] -a <ampl.fa> -s [<in.sam>] [-o <out.sam>] [-e <err.sam>]
```

.fa indicates fasta format

.fq indicates fastq format

.sam indicates sam format

```
<ampl.fa>    Amplicons
<in.*>      File with unpaired reads (default: stdin)
<in1.*>     File with #1 mates
<in2.*>     File with #2 mates

<out.*>     Trimmed unpaired reads (default: stdout)
<out1.*>    Trimmed #1 mates
<out2.*>    Trimmed #2 mates
<err.*>     Unmatched unpaired reads (default: <out.*>)
<err1.*>    Unmatched #1 mates (default: <out1.*>)
<err2.*>    Unmatched #2 mates (default: <out2.*>)
```

Options:

```
-k INT      K-mer size (default: 13)
-t FLOAT    Alignment score ratio threshold (default: 0.5)
-m FLOAT    Mismatch penalty (default: -1.4)
-g FLOAT    Gap penalty (default: -1.4)
-d INT      Max number of degenerate nucleotides N in a k-mer (default: 2)
-b          Trim paired reads from both sides
```

Other:

```
-h/--help    Show this message
```

Conclusion

After using prettry both Samtools and GATK detected **E92K** in all 9 patients

| **Questions?**