Bring Flow Cells!

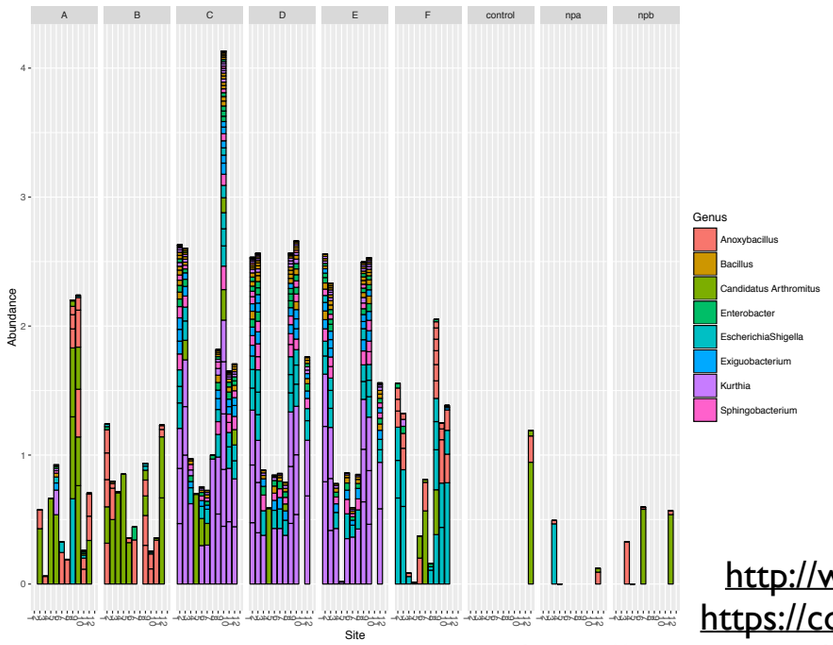
HTS Background and Theory

Josh Granek

Molecular Biology







http://www.geograph.org.uk/photo/2847164
https://commons.wikimedia.org/wiki/File:Pcr.jpg
https://commons.wikimedia.org/wiki/File:Illumina_MiSeq_sequencer.jpg

Bioinformatic Analysis

```
@M00698:36:000000000-AFBEL:1:1101:14738:1412 1:N:0:0
TTACGCTAACAGGCGGTAGCCTGGCAGGGTCAGGAAATCAATTAACTCATCGGAAGTGGTGATCTGTTCCATCAAGCGTGCGGCATCGTCAAAACGCCC
+
ABBBABBBAFFFGGGGGGGGGGHGGHGGGCG2GF3FFGHHHHHHHGGFGHEHHGGGEHHHHAGGHHGHHHFFDHFHHHGEGGGG@F@H?GHH/GBEFGGG
@M00698:36:000000000-AFBEL:1:1101:16483:1412 1:N:0:0
CTGCCAGTTGAACGACGGCGAGCAGTTATAAGCCAGCAGTTTGCCCGGATATTTCGCGTGGATAGCTTGTGCAAAGCGACGCGCCAGTTCCAGATCCGGCG
+
AAABBFFFFFFFGGGGGGGGGGGGHHHHHHHHHGHGHGHHHHHGHHHGGGGGHHHHGGGGGGGHHHGHHFFHHHHHGHGGGGGGGGGGHHHHHHHHHHHGGG
@M00698:36:000000000-AFBEL:1:1101:15928:1413 1:N:0:0
GTAAAGTCCTGAGTGATACCGGCAACTTTTACCCCCAGTCCCACTTTCGAACCGGCAAACATATCGGCAAAAGAGGCCGTGCCTGATTTAAAGCCGTAGGT
+
```

	Sample 1	Sample 2	…	Sample N
Bacteria 1				
Bacteria 2				
…				
Bacteria N				

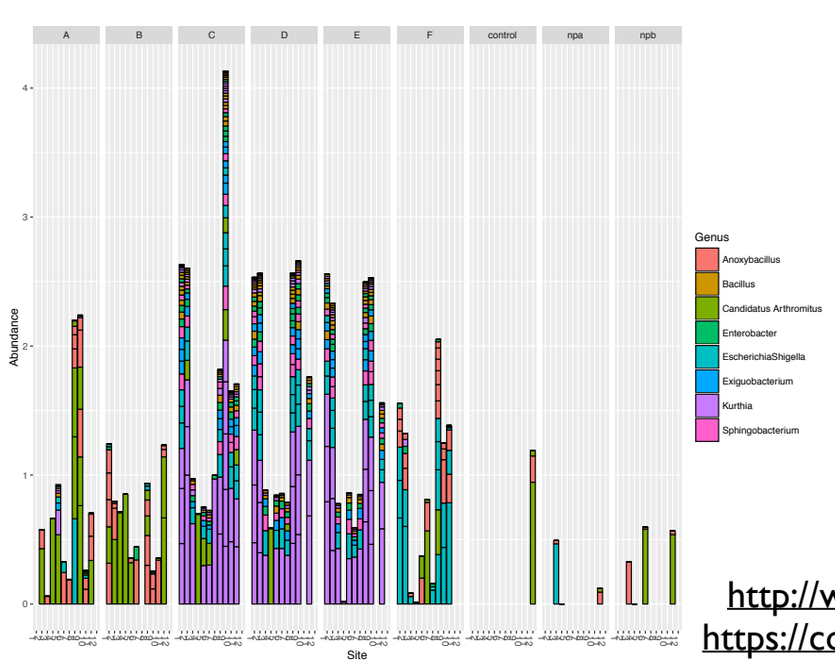
https://commons.wikimedia.org/wiki/File:Illumina_MiSeq_sequencer.jpg

Statistical Analysis

	Sample 1	Sample 2	…	Sample N
Bacteria 1				
Bacteria 2				
…				
Bacteria N				

1. What is present?
2. How much?
3. Are there differences between treatments, host species, …?
4. What are the differences?

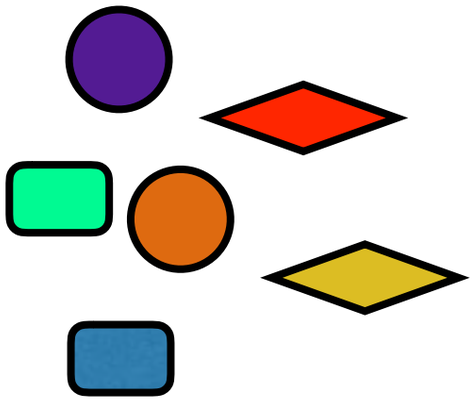
Molecular Biology



http://www.geograph.org.uk/photo/2847164
https://commons.wikimedia.org/wiki/File:Pcr.jpg
https://commons.wikimedia.org/wiki/File:Illumina_MiSeq_sequencer.jpg

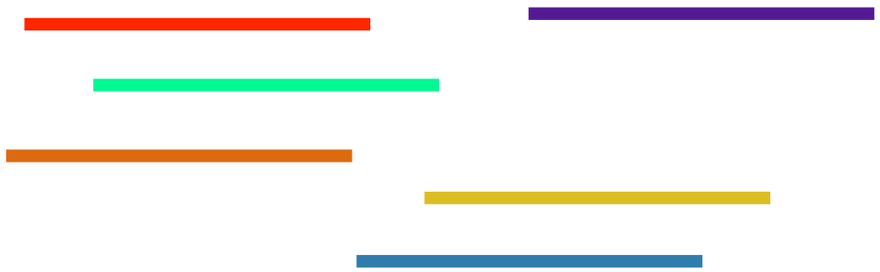
Molecular Biology

Microbes



DNA Purification

DNA

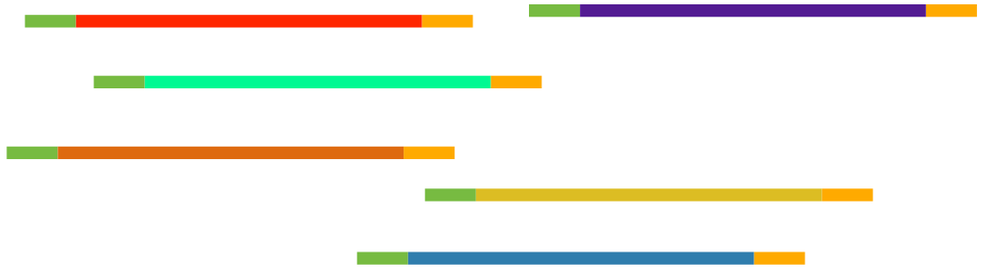


\+

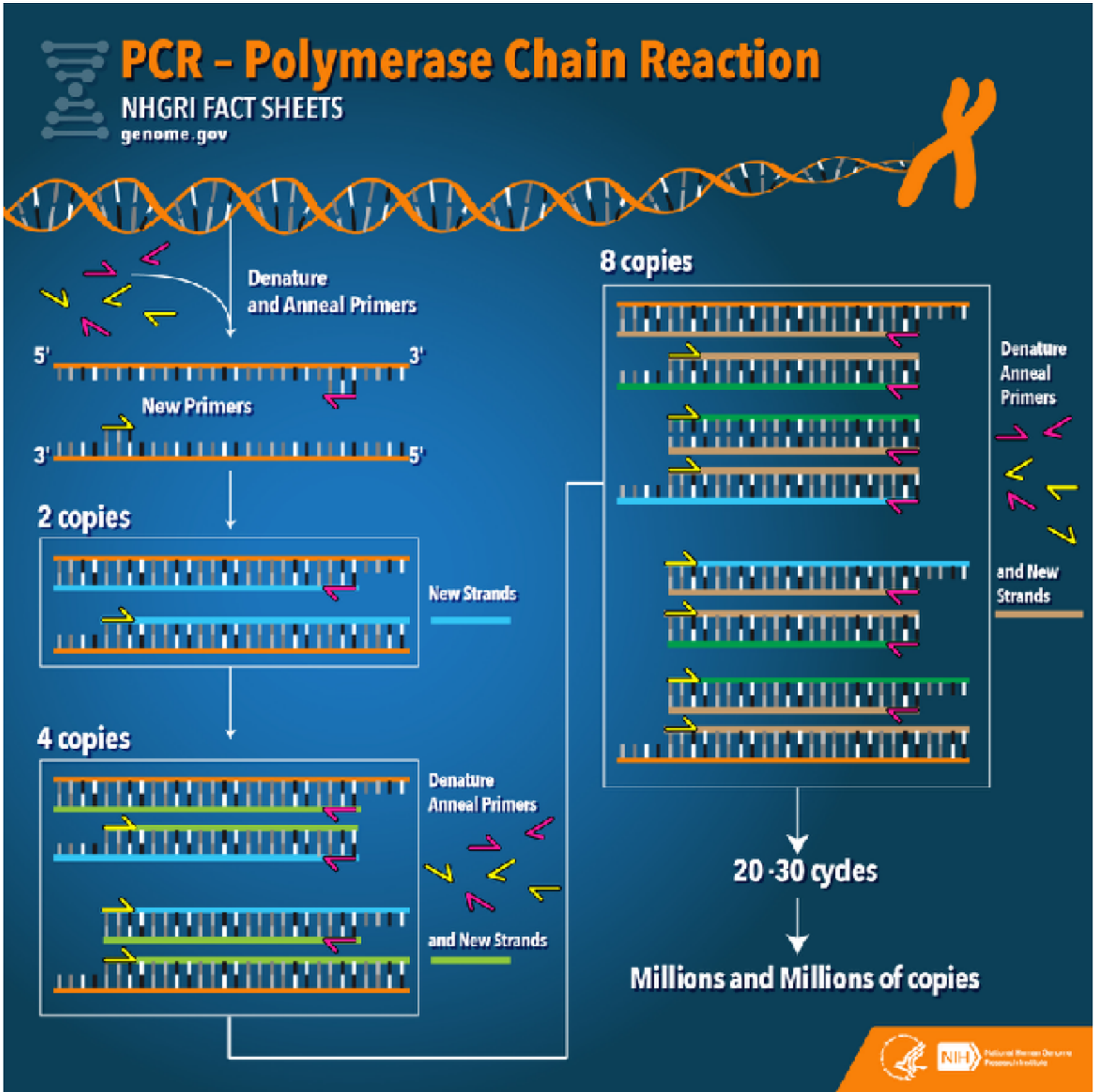
Primers

Library Preparation

Amplicon Library



PCR



https://www.genome.gov/images/content/pcr_factsheet.jpg

Sanger Sequencing

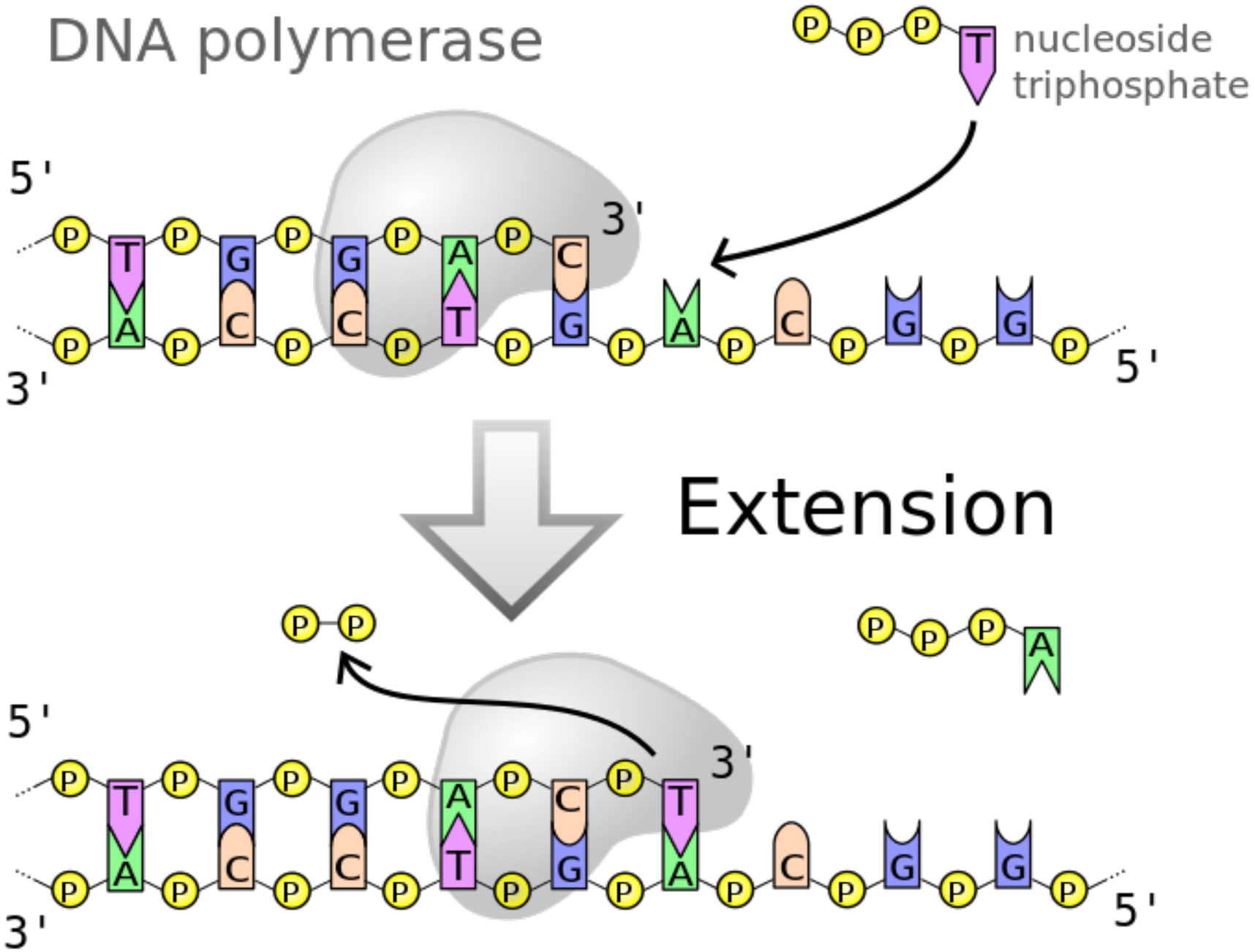
DNA Synthesis

- What are the minimum components for DNA Replication?

DNA Synthesis

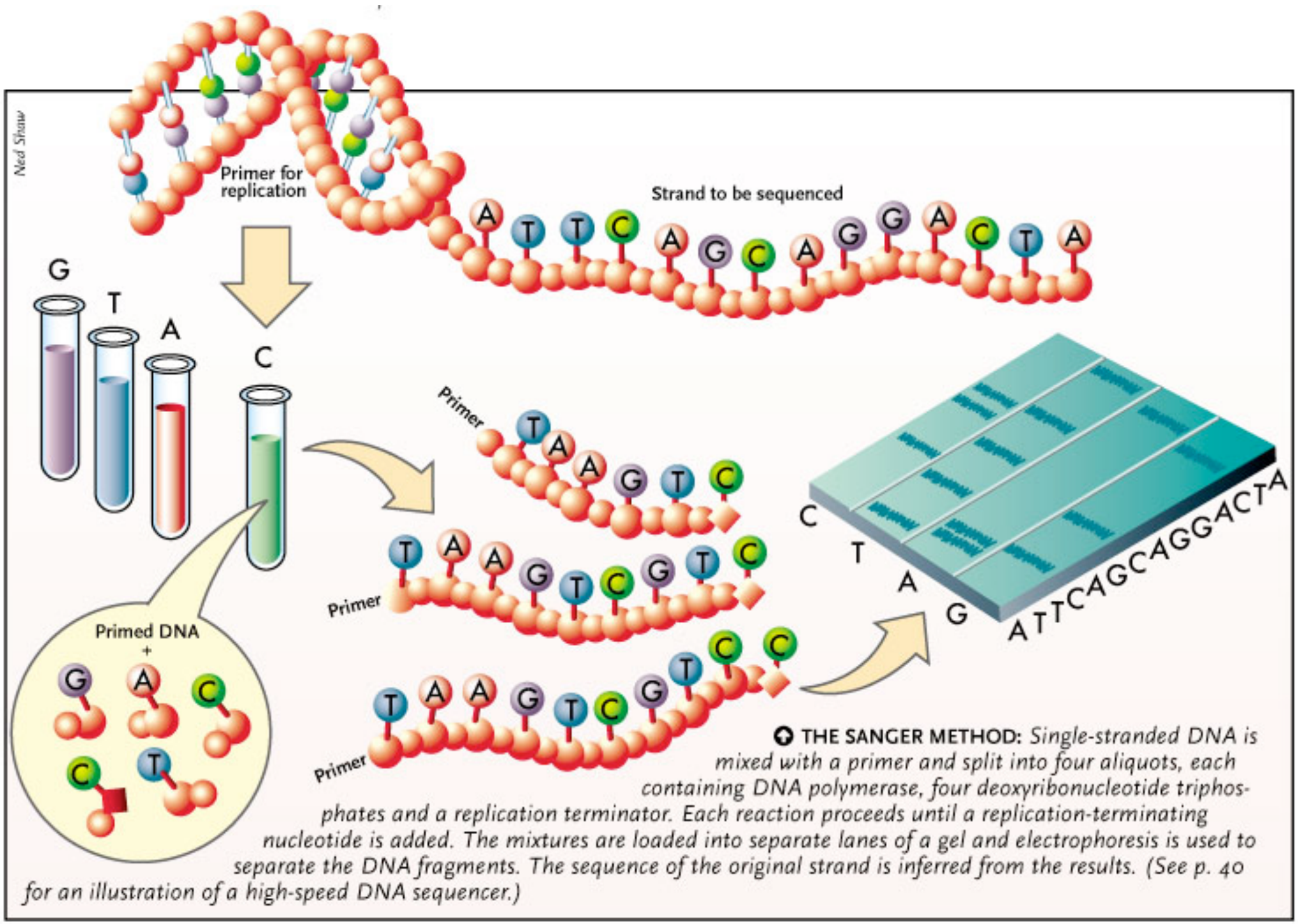
- What are the minimum components for DNA Replication?
  - Template
  - Primer
  - Nucleoside triphosphates
  - DNA Polymerase*

DNA Synthesis



https://upload.wikimedia.org/wikipedia/commons/thumb/6/6f/DNA_polymerase.svg/700px-DNA_polymerase.svg.png

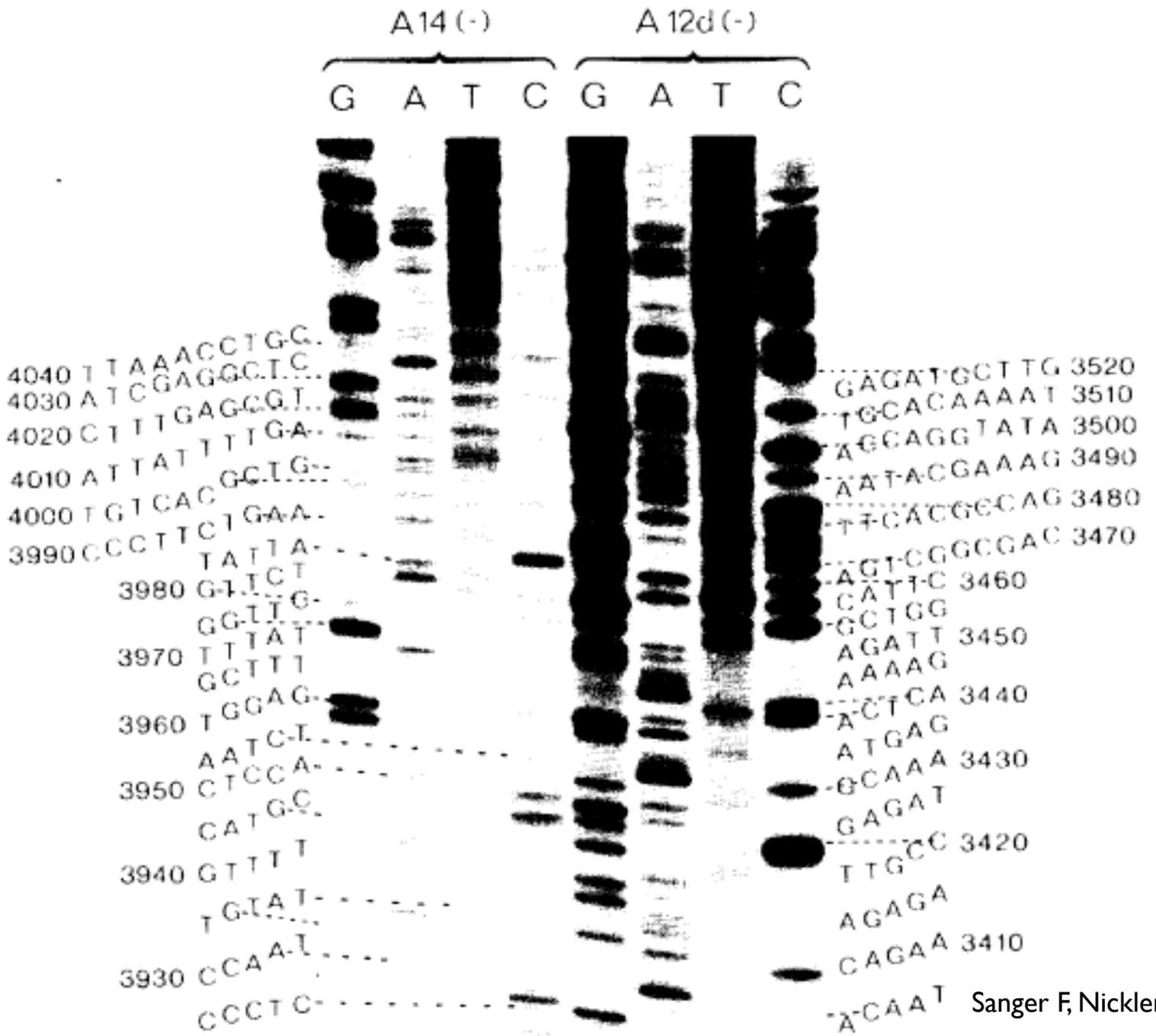
Sanger Sequencing



THE SANGER METHOD: *Single-stranded DNA is mixed with a primer and split into four aliquots, each containing DNA polymerase, four deoxyribonucleotide triphosphates and a replication terminator. Each reaction proceeds until a replication-terminating nucleotide is added. The mixtures are loaded into separate lanes of a gel and electrophoresis is used to separate the DNA fragments. The sequence of the original strand is inferred from the results. (See p. 40 for an illustration of a high-speed DNA sequencer.)*

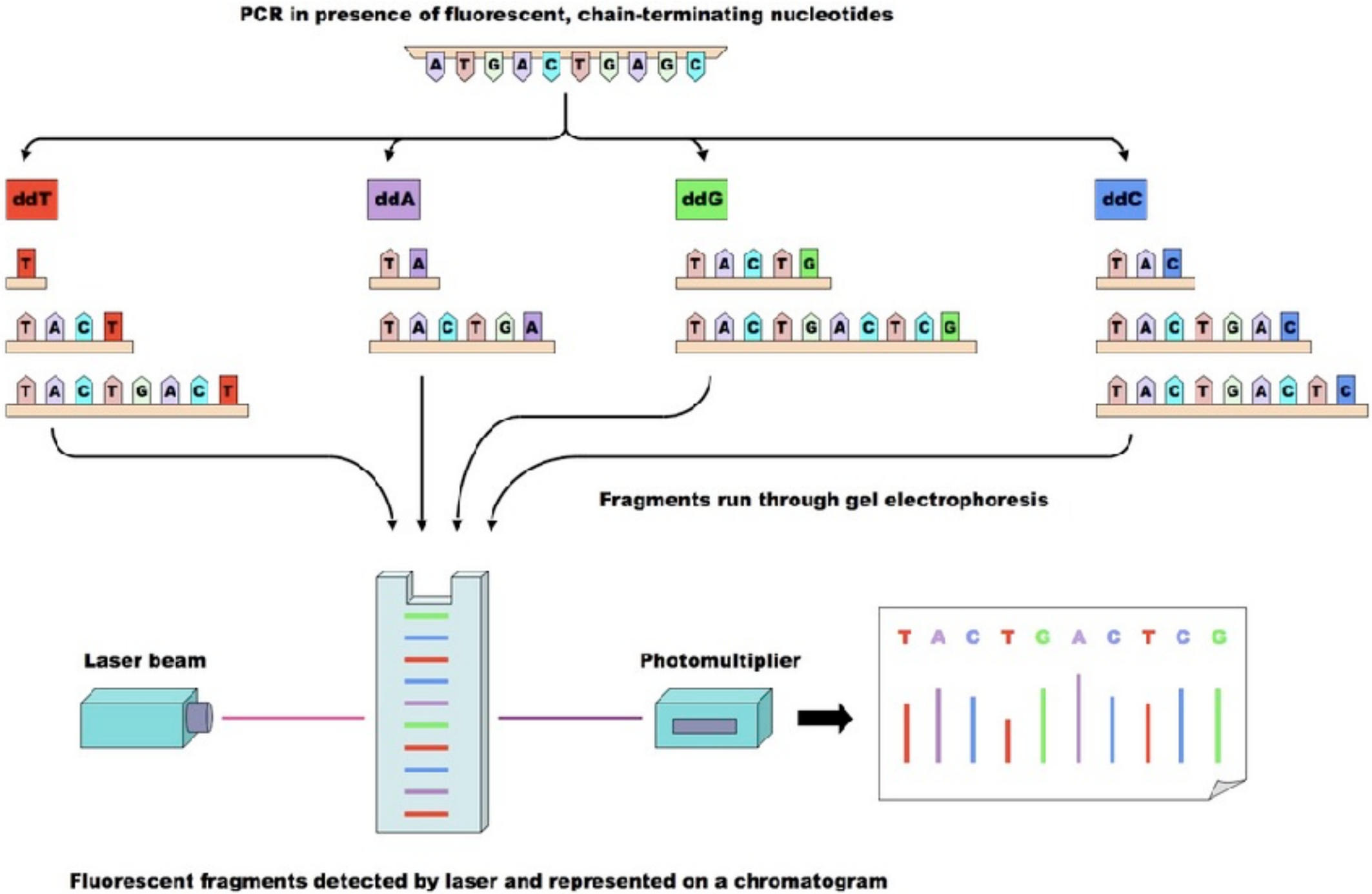
http://www.the-scientist.com/?articles.view/articleNo/15939/title/DNA-Sequencing-Industry-Sets-its-Sights-on-the-Future/

Sanger Sequencing



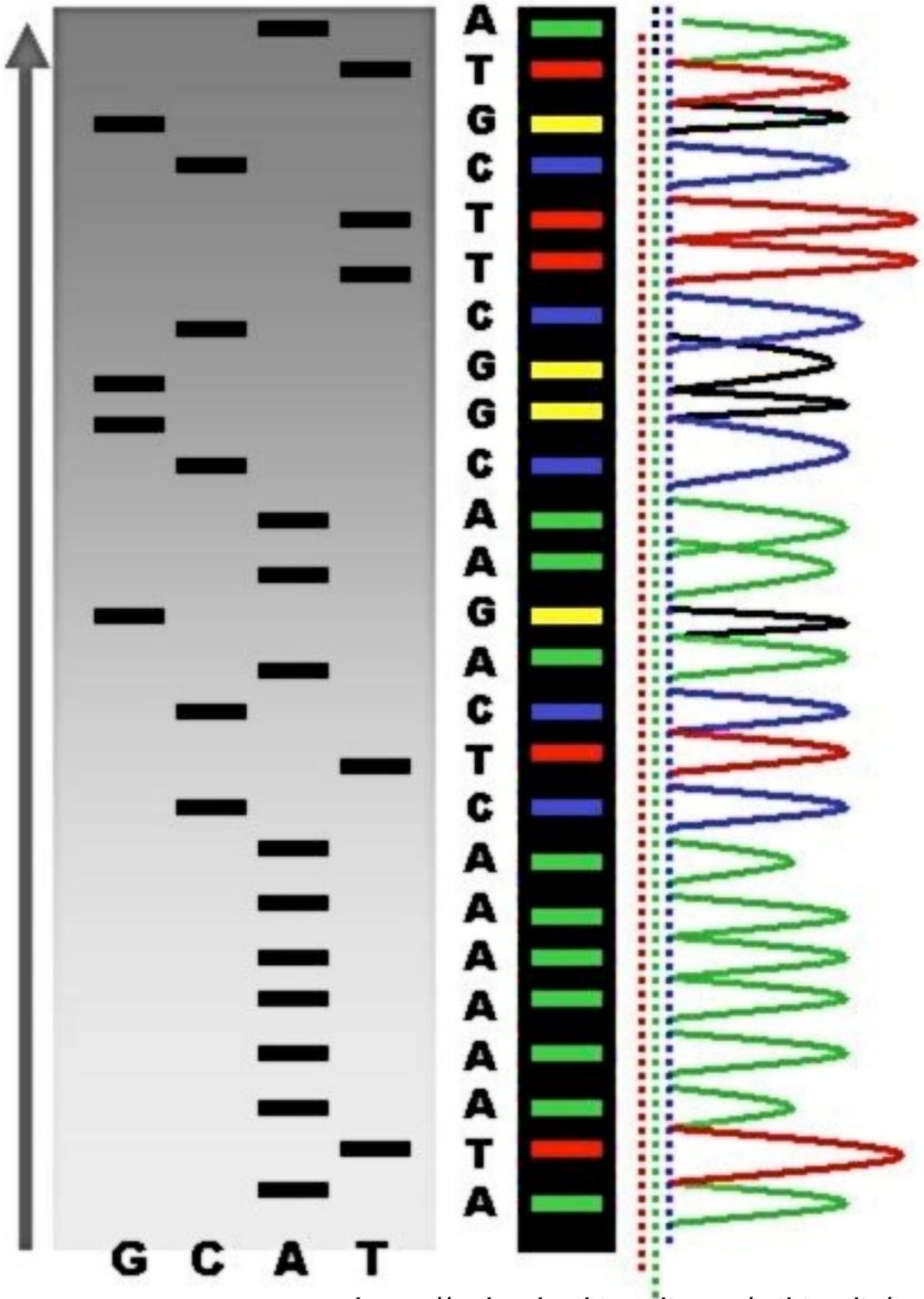
Sanger F, Nicklen S, Coulson AR., *PNAS*, 1977

Dye-terminator



http://www.vce.bioninja.com.au/_Media/sequencing_med.jpeg

Radiolabel vs. Dye



https://upload.wikimedia.org/wikipedia/commons/3/3d/Radioactive_Fluorescent_Seq.jpg

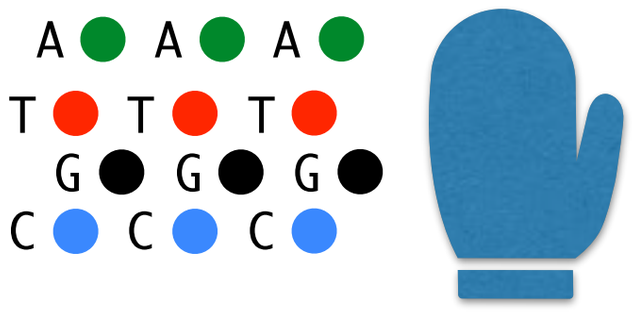
High-Throughput Sequencing

Sequencing

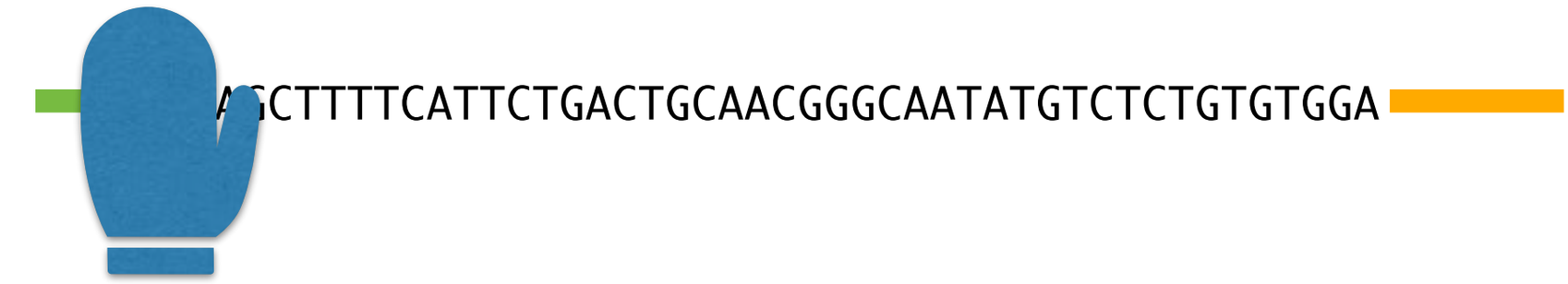
Sequencing

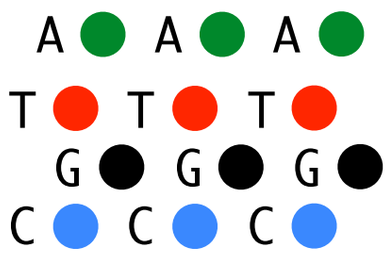
Sequencing


AGCTTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGA

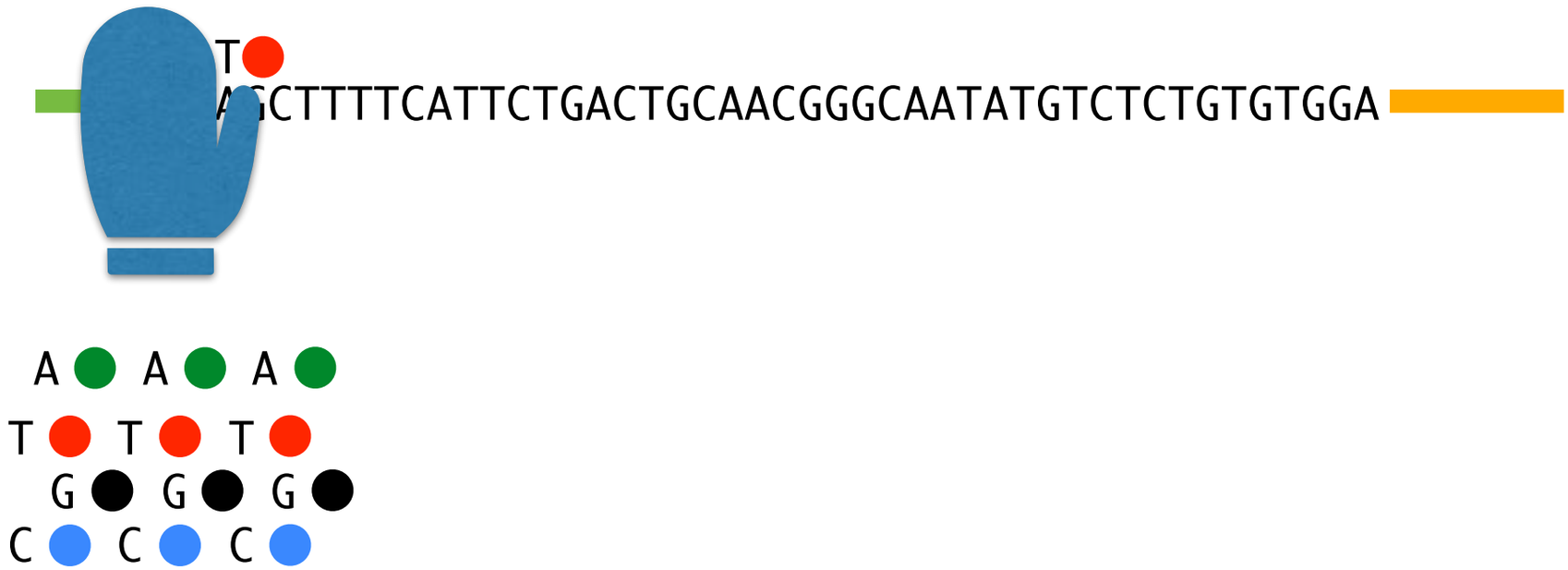

A A A
T T T
G G G
C C C

Sequencing


AGCTTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGA


A A A
T T T
G G G
C C C

Sequencing

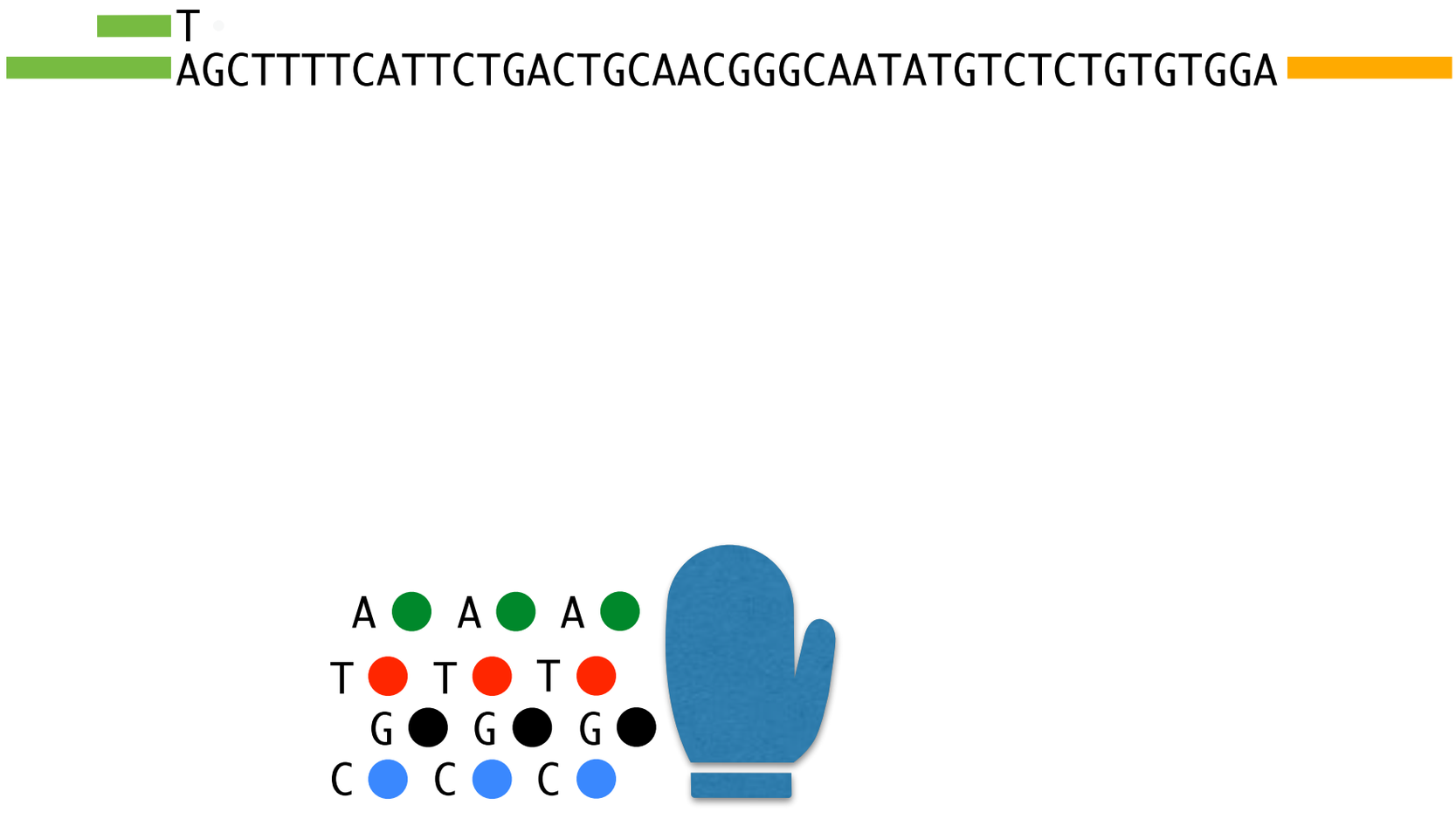

T
AGCTTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGA
A A A
T T T
G G G
C C C

Sequencing

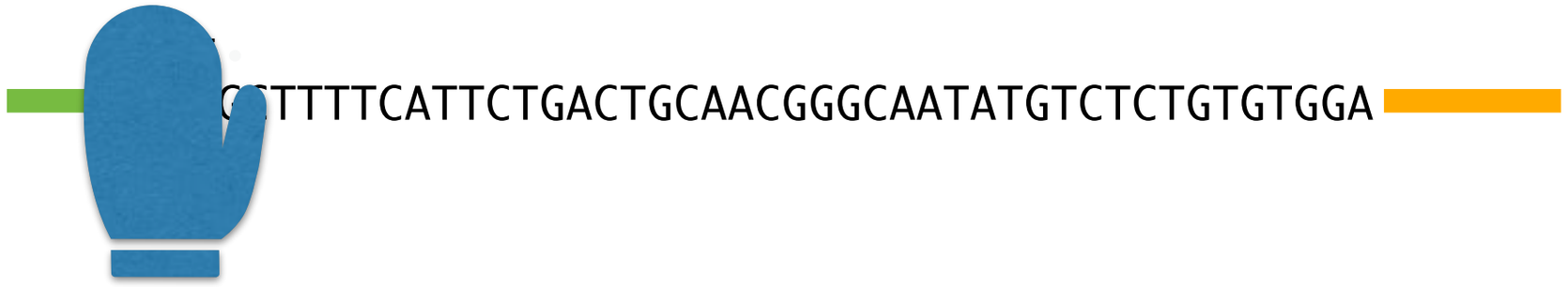
Sequencing

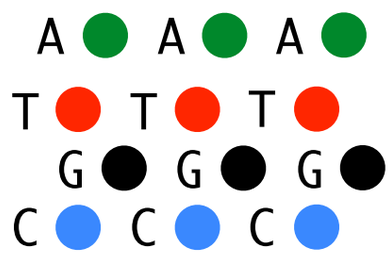
Sequencing

Sequencing

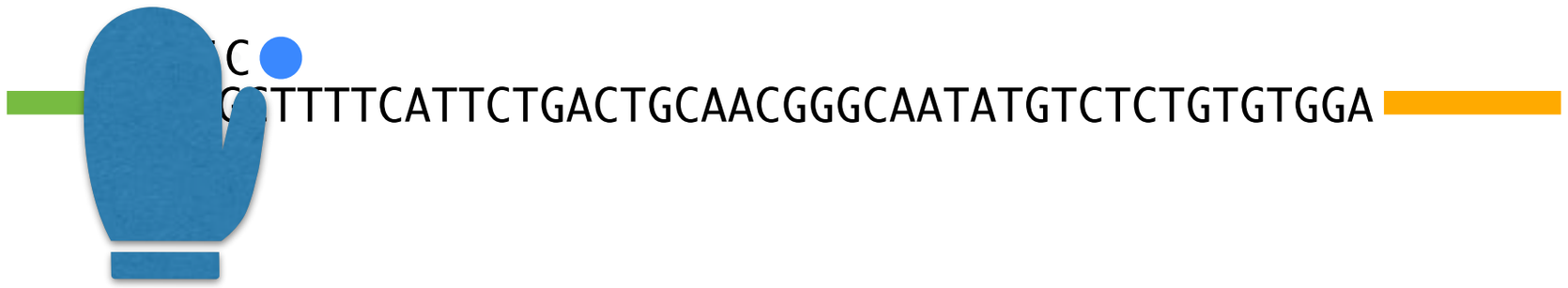

T
AGCTTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGA
A A A
T T T
G G G
C C C

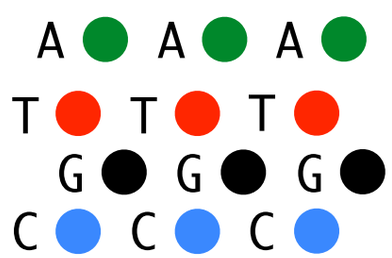
Sequencing


CTTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGA


A A A
T T T
G G G
C C C

Sequencing


C
CTTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGA

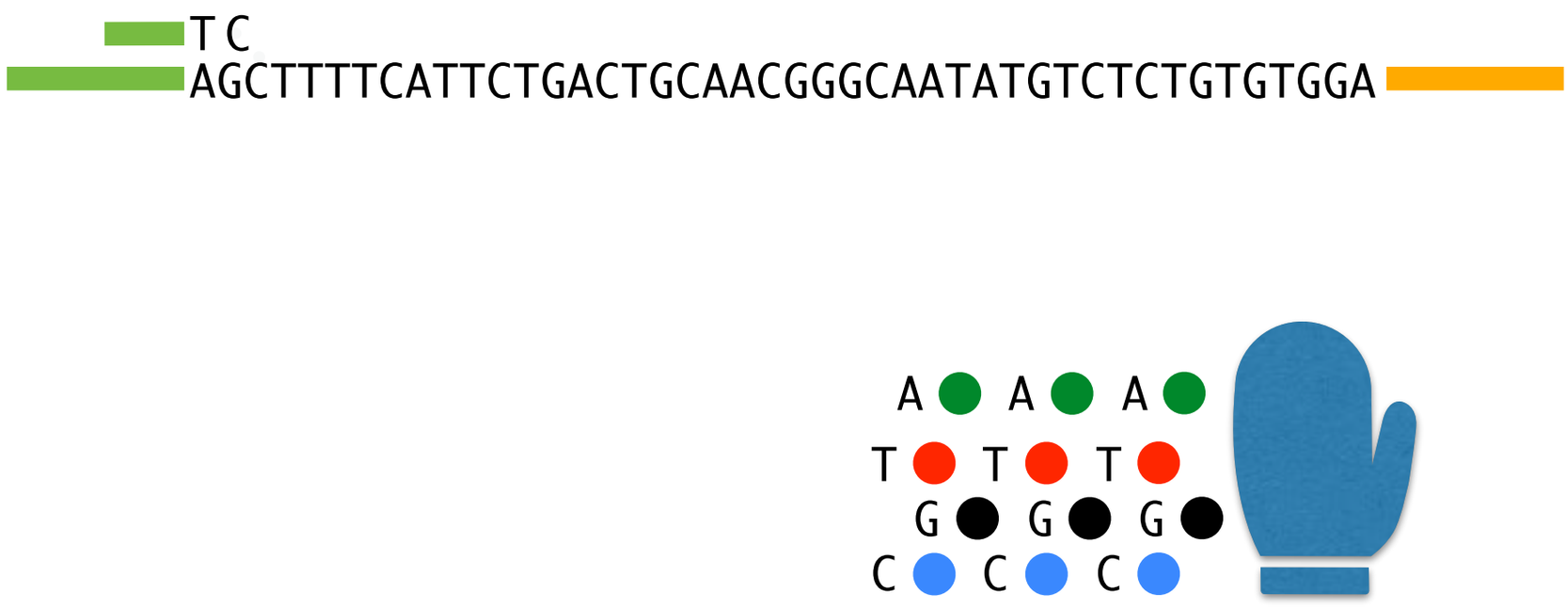

A A A
T T T
G G G
C C C

Sequencing

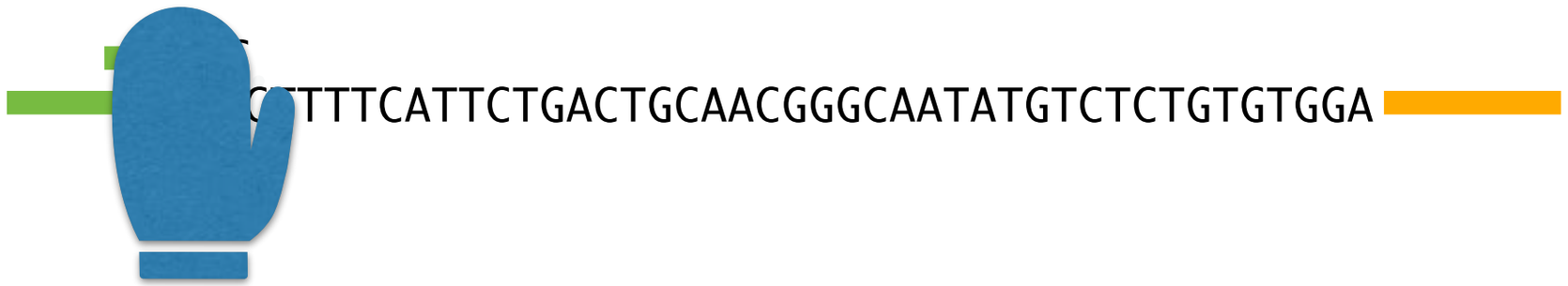
Sequencing

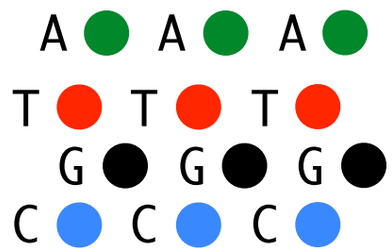
Sequencing

Sequencing

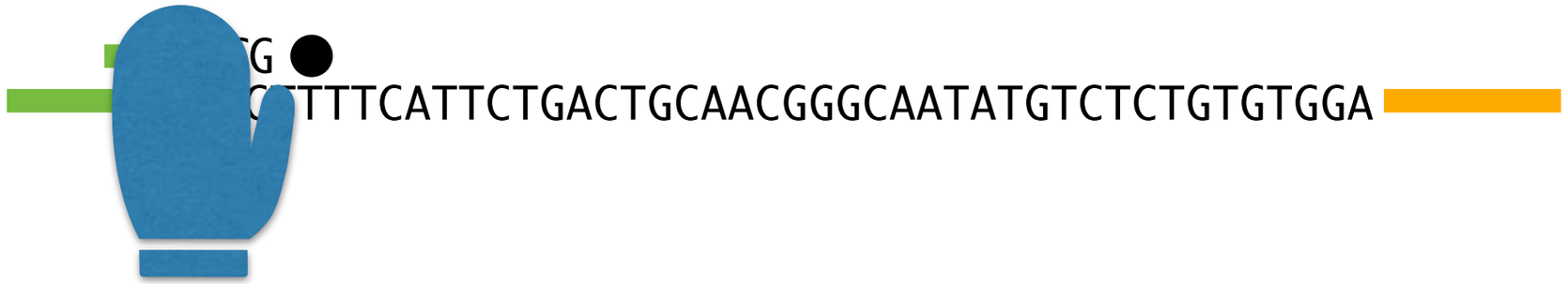

T C
AGCTTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGA
A A A
T T T
G G G
C C C

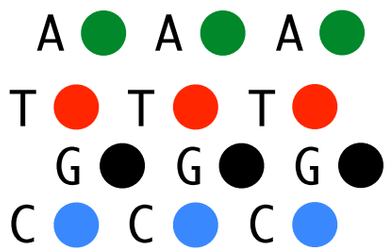
Sequencing


TTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGA


A A A
T T T
G G G
C C C

Sequencing

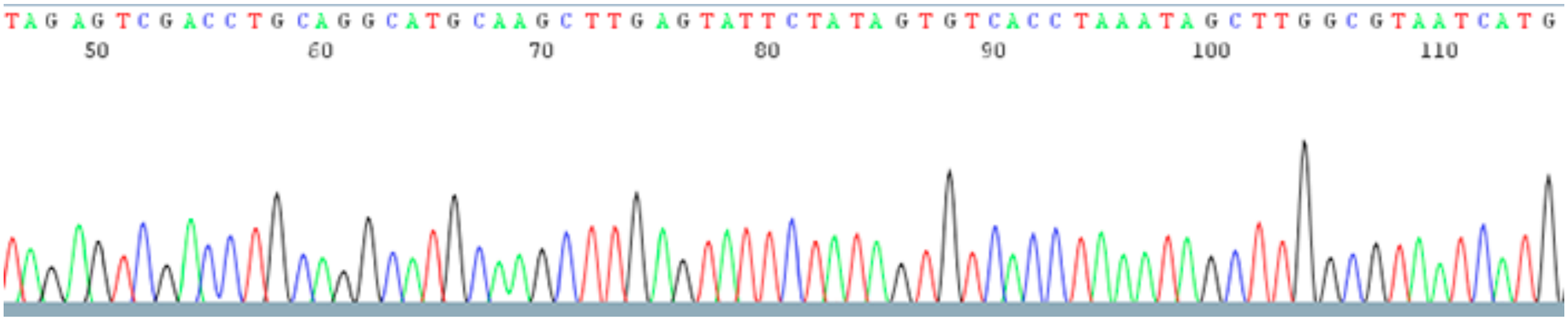

G
TTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGA


A A A
T T T
G G G
C C C

Sequencing

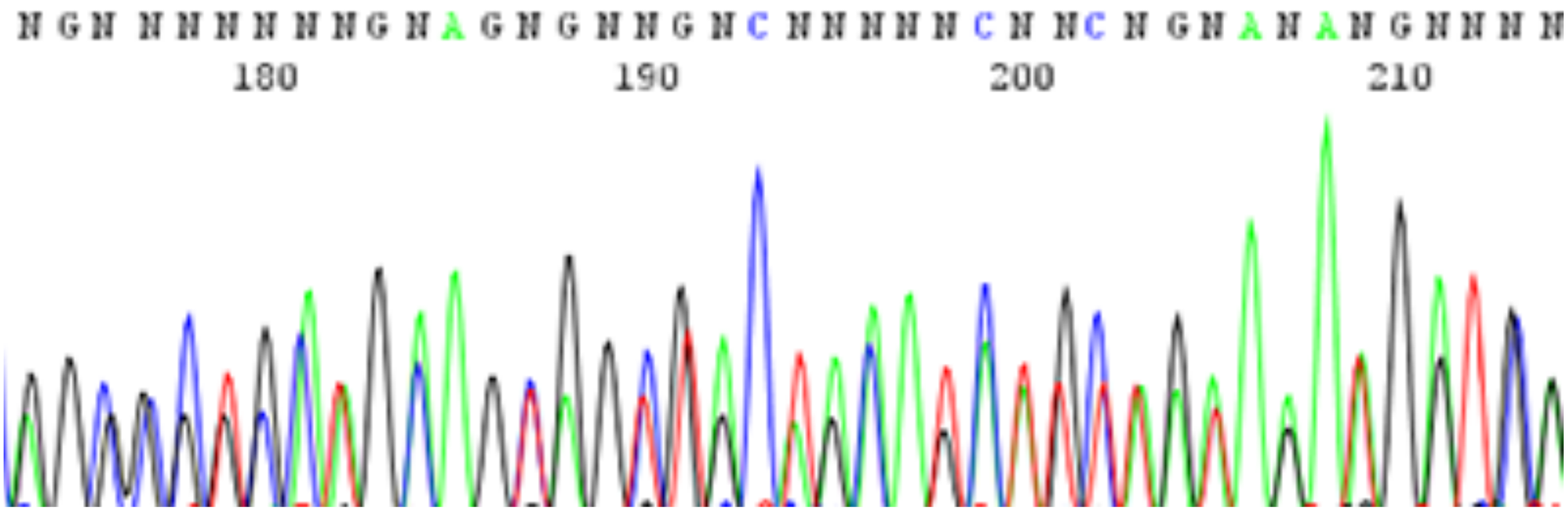
Sequencing

Dye-terminator Sanger Sequencing



https://dnacore.mgh.harvard.edu/new-cgi-bin/site/pages/sequencing_pages/seq_troubleshooting.jsp

Double Sequence



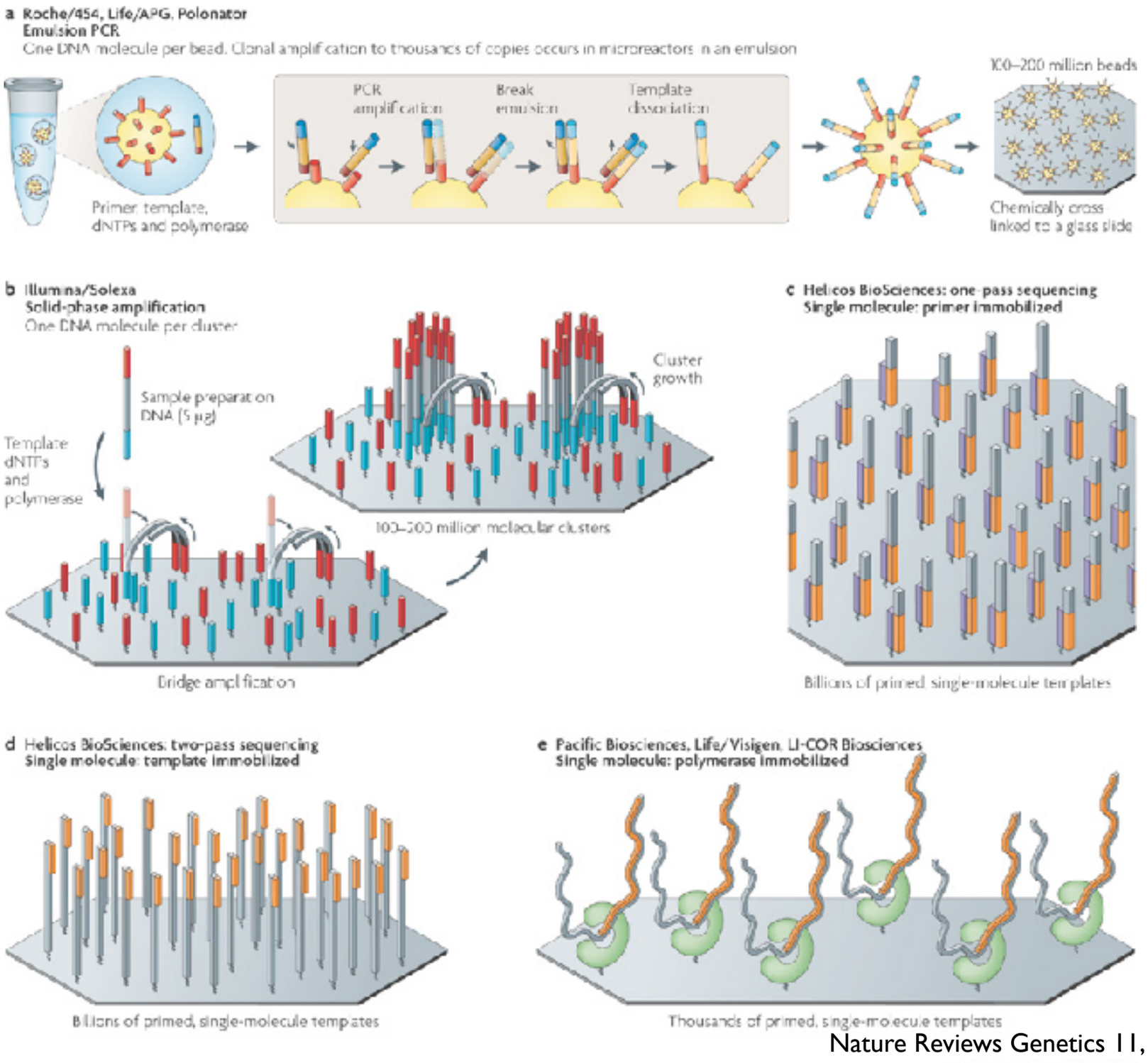
https://dnacore.mgh.harvard.edu/new-cgi-bin/site/pages/sequencing_pages/seq_troubleshooting.jsp

How?

How?

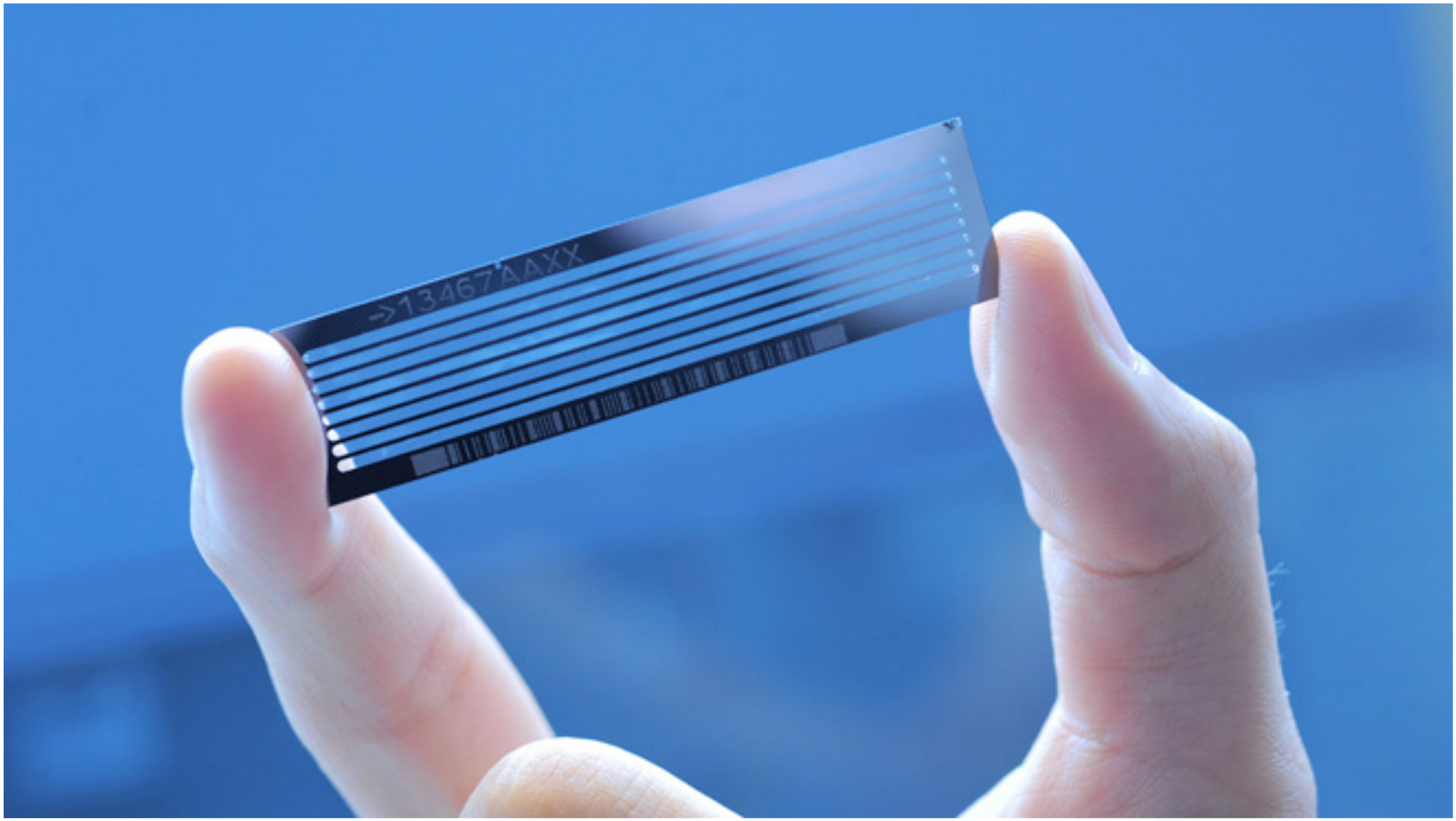
- Separate
- Detect
- Removable Terminator

Template immobilization



Nature Reviews Genetics 11, 31-46, doi:10.1038/nrg2626

A Flow Cell



http://global.fncstatic.com/static/managed/img/Scitech/neanderthal%20genome%205.jpg

Pass Around Flow Cells!!!

SBS: Sequencing by Synthesis

An Illumina Story

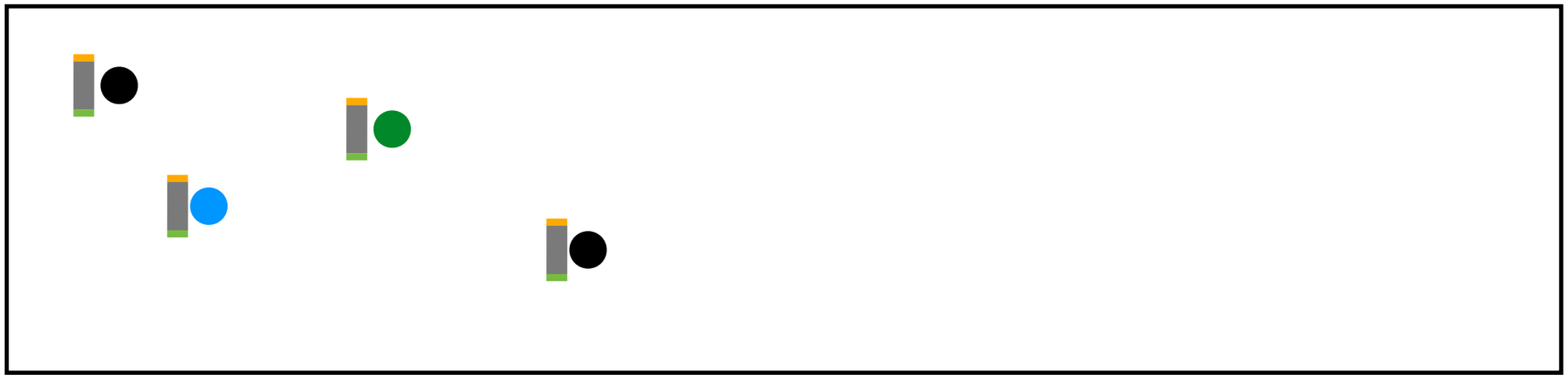
A Flow Cell

Bind Library

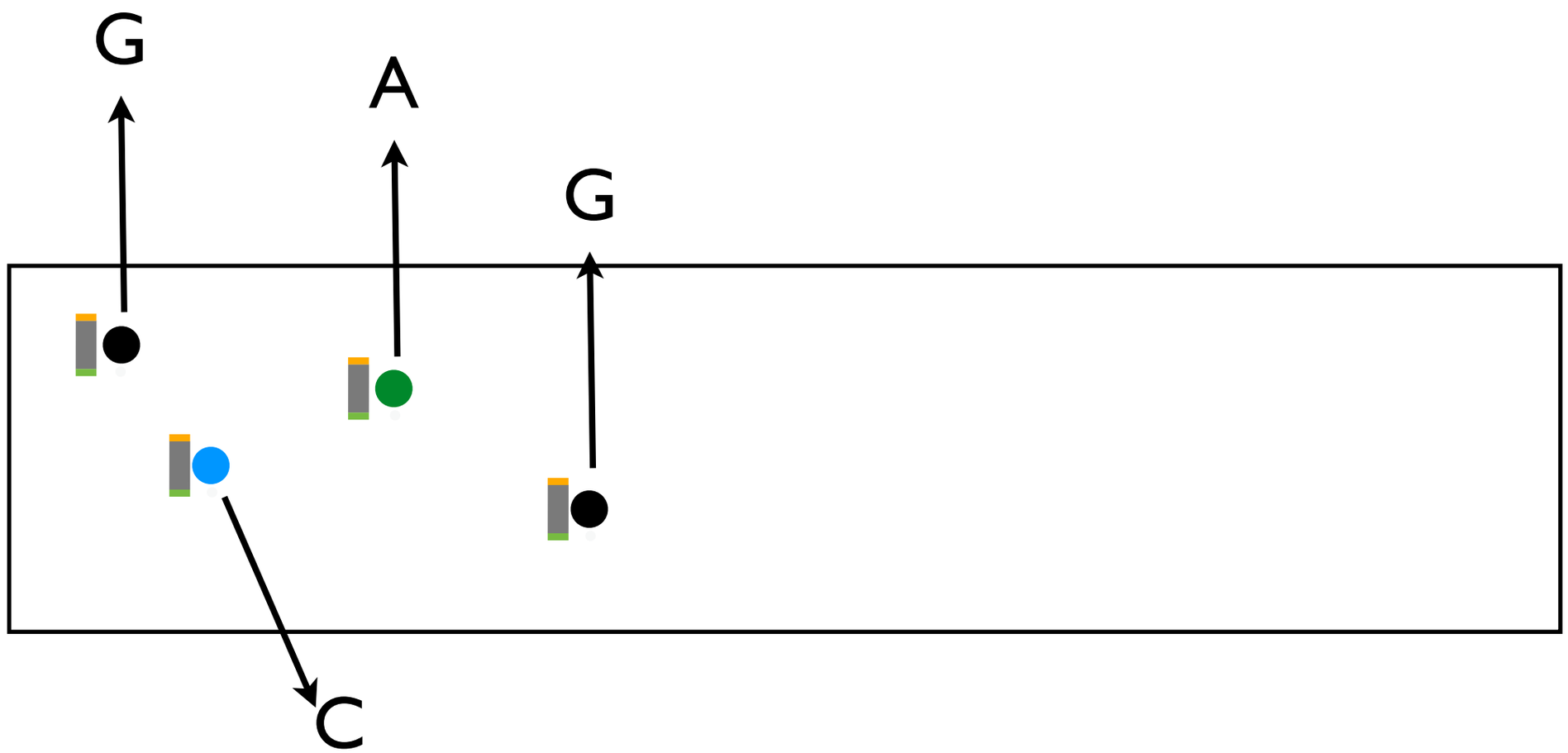
1st Cycle

1st Cycle

1st Cycle



1st Cycle

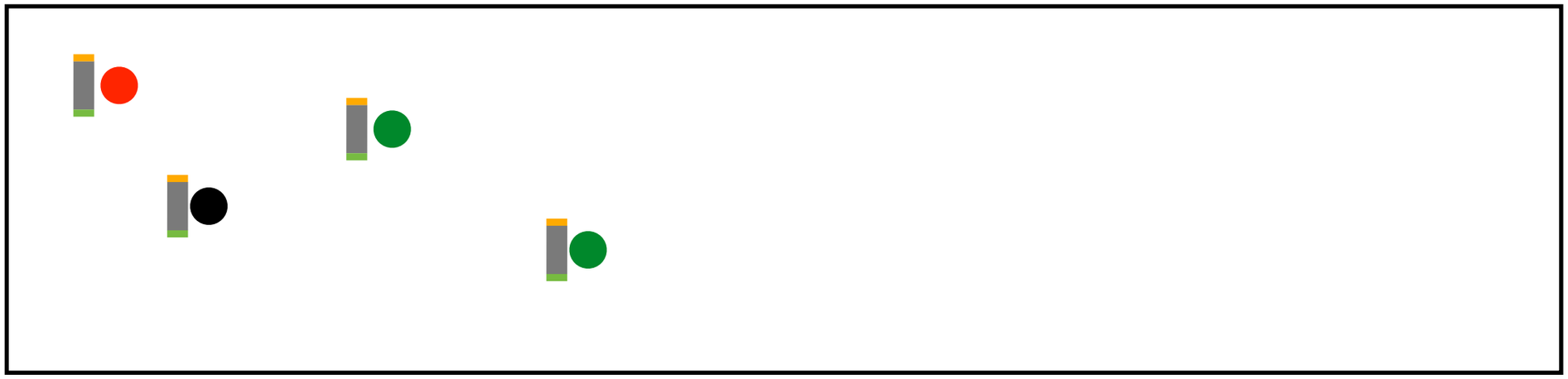


2nd Cycle

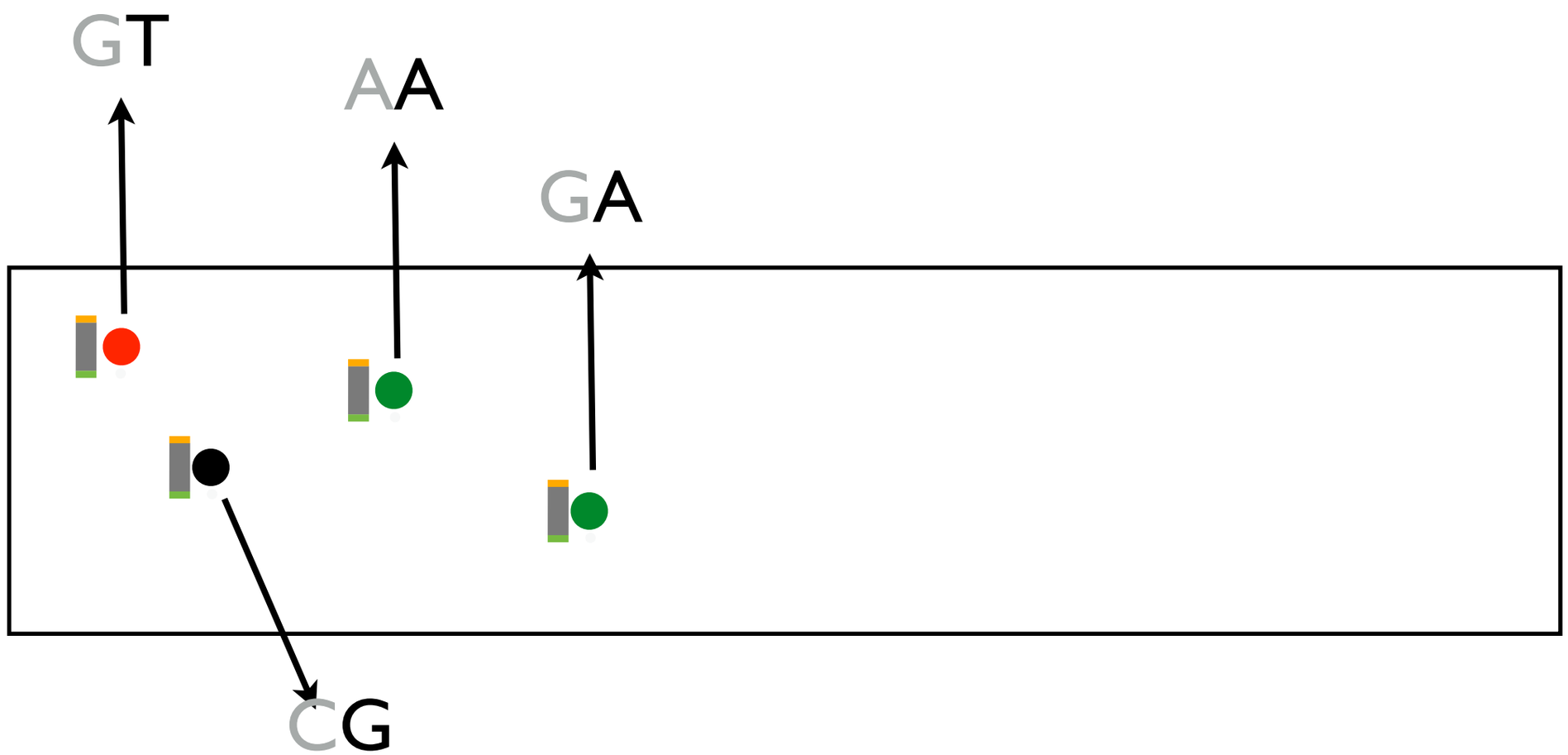
2nd Cycle

2nd Cycle

2nd Cycle



2nd Cycle

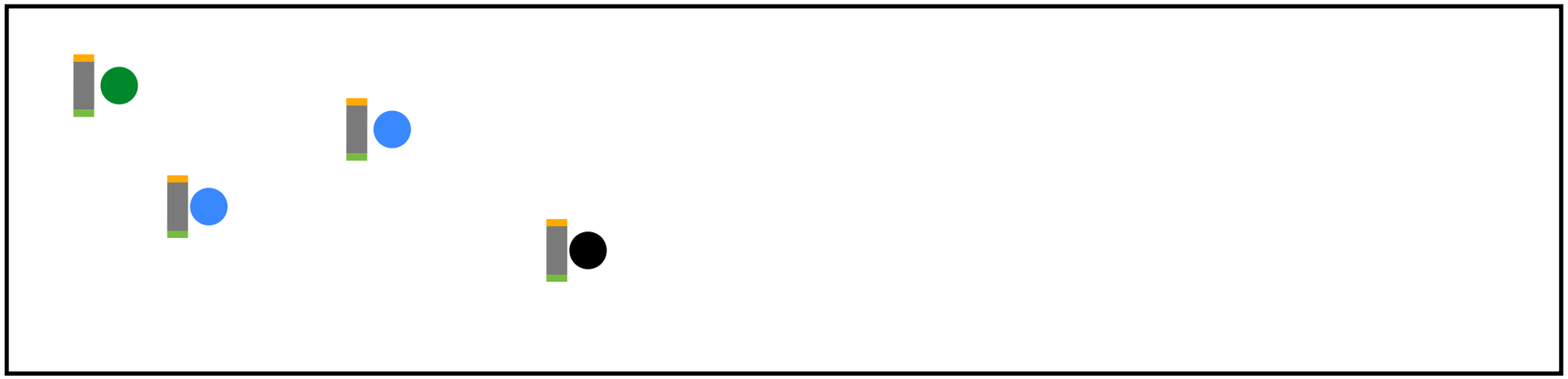


3rd Cycle

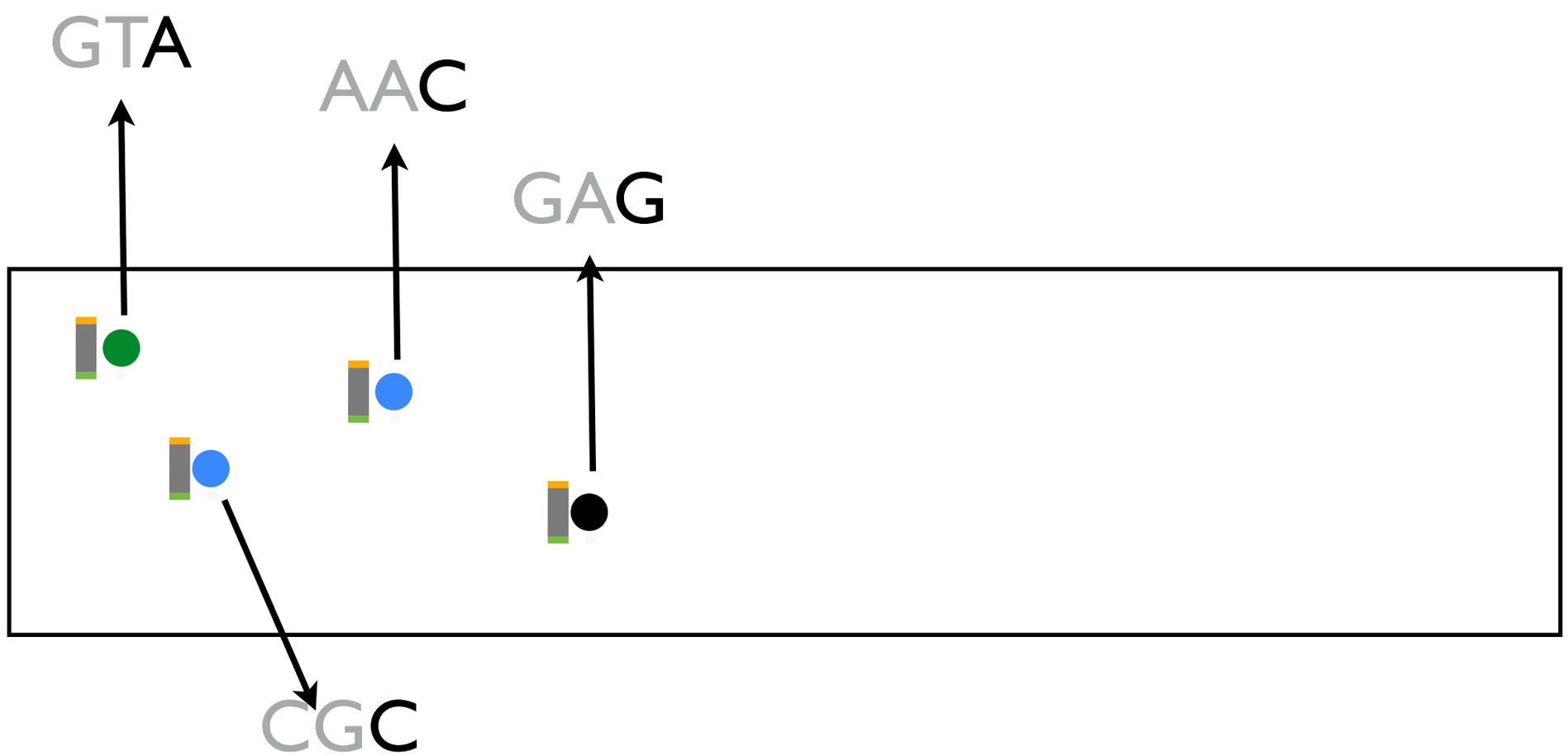
3rd Cycle

3rd Cycle

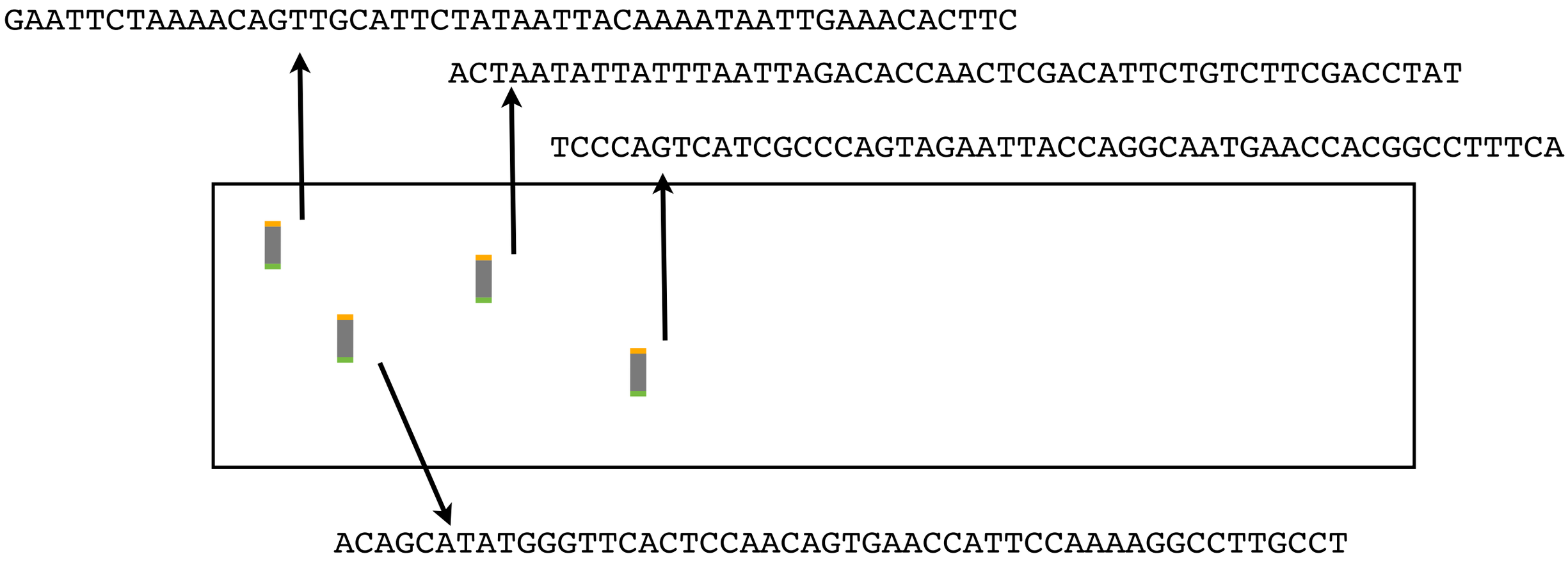
3rd Cycle



3rd Cycle



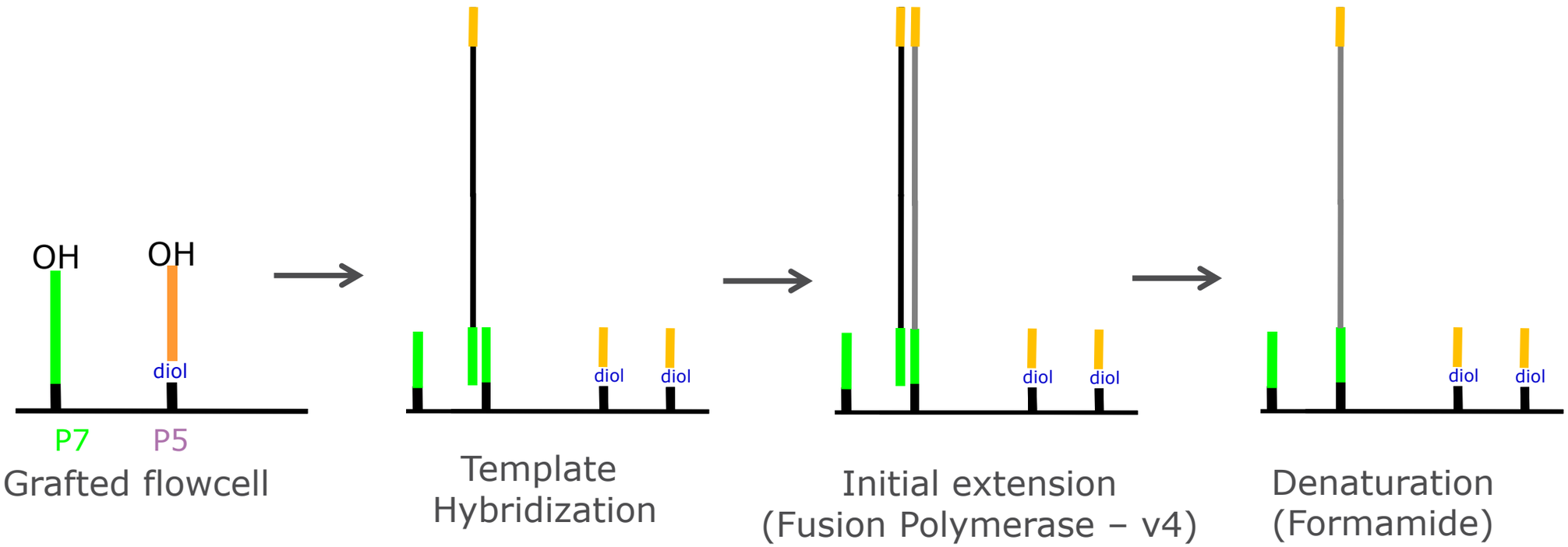
50th Cycle


GAATTCTAAAACAGTTGCATTCTATAATTACAAAATAATTGAAACACTTC
ACTAATATTATTTAATTAGACACCAACTCGACATTCTGTCTTCGACCTAT
TCCCAGTCATCGCCCAGTAGAATTACCAGGCAATGAACCACGGCCTTTCA
ACAGCATATGGGTTCACTCCAACAGTGAACCATTCCAAAAGGCCTTGCCT

Illumina Short Reads

- 50 - 300bp

Cluster generation – hybridization and amplification



Hybridization

```
5′-CTGATCTGACTGATGCGTATGCTAGT-3'
```

+

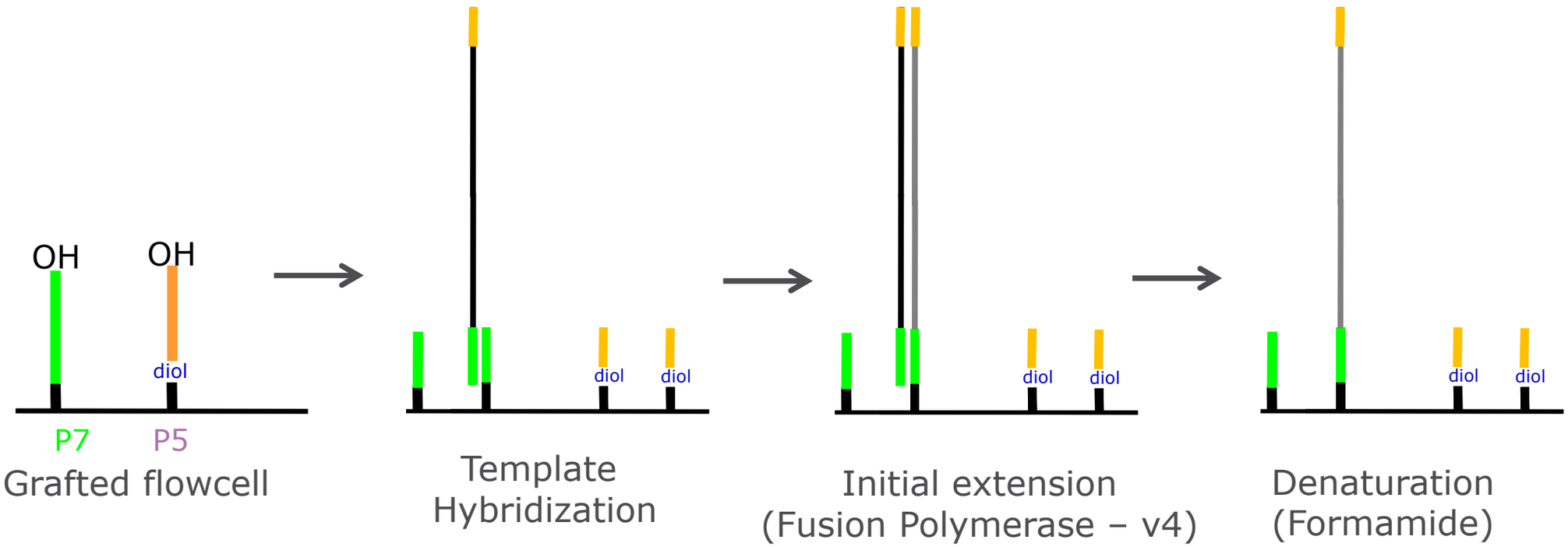
```
3′-GCATAC-5′
```

=

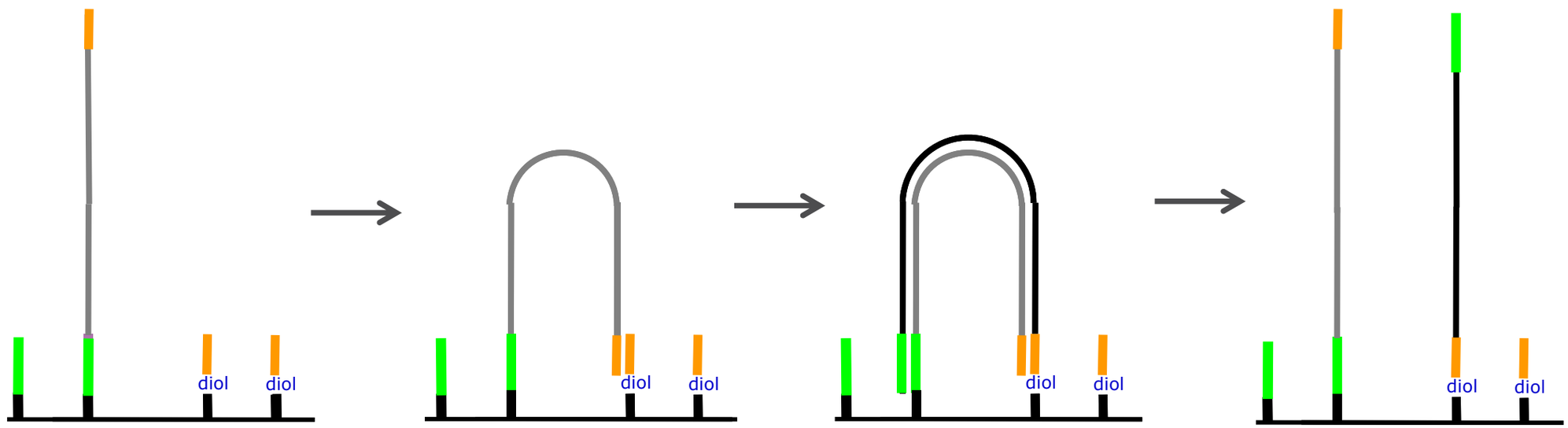
```
5′-CTGATCTGACTGATGCGTATGCTAGT-3'
      3′-GCATAC-5′
```

Derived from:
https://en.wikipedia.org/wiki/Sticky_and_blunt_ends

Cluster generation – hybridization and amplification

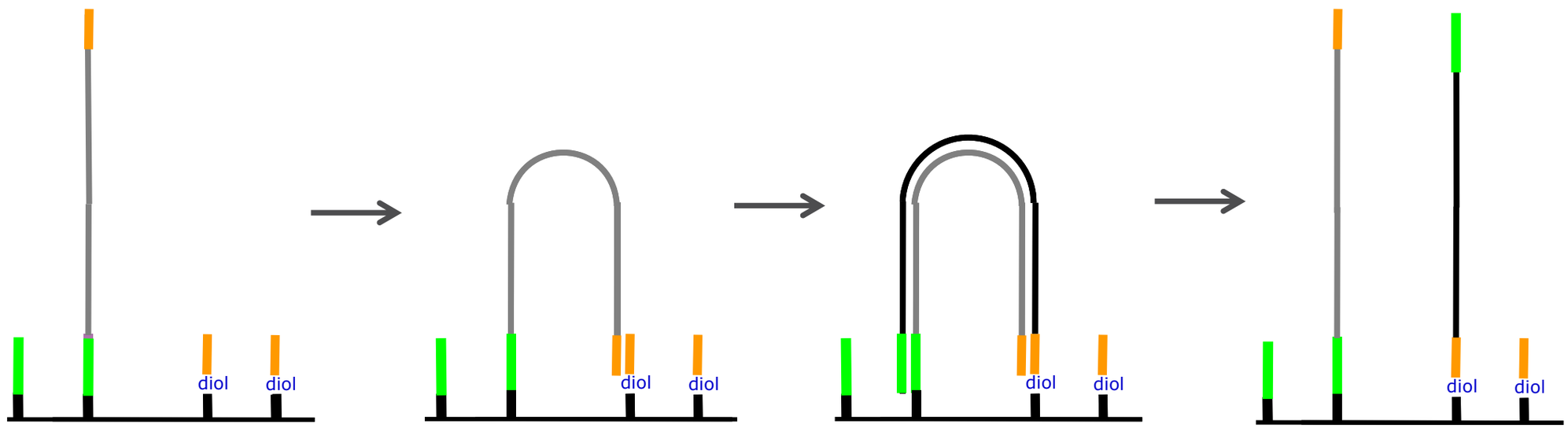


Cluster generation – hybridization and amplification



illumina

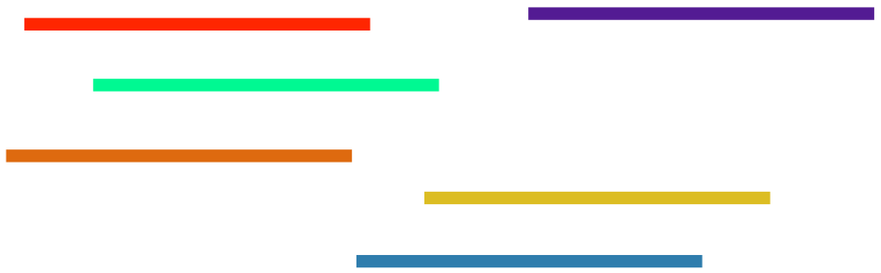
Cluster generation – hybridization and amplification



illumina

Library Preparation

DNA

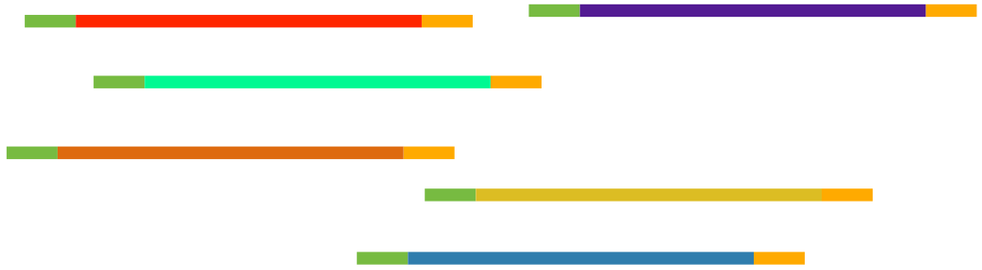


+

Primers

Library Preparation

Amplicon Library



Why Adapters?

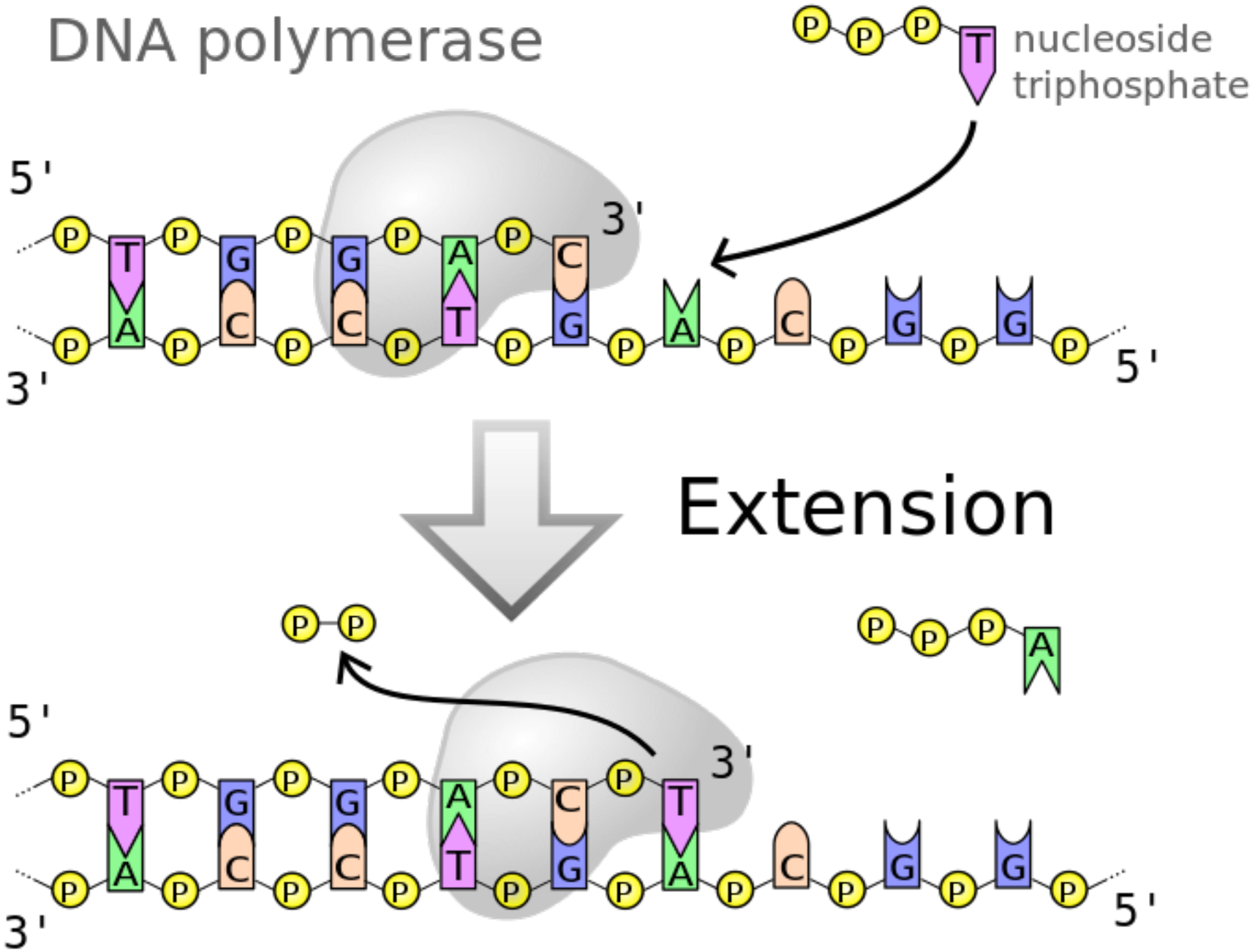
DNA Synthesis

- What are the minimum components for DNA Replication?

DNA Synthesis

- What are the minimum components for DNA Replication?
  - Template
  - Primer
  - Nucleoside triphosphates
  - DNA Polymerase*

DNA Synthesis



https://upload.wikimedia.org/wikipedia/commons/thumb/6/6f/DNA_polymerase.svg/700px-DNA_polymerase.svg.png

Why Adapters?

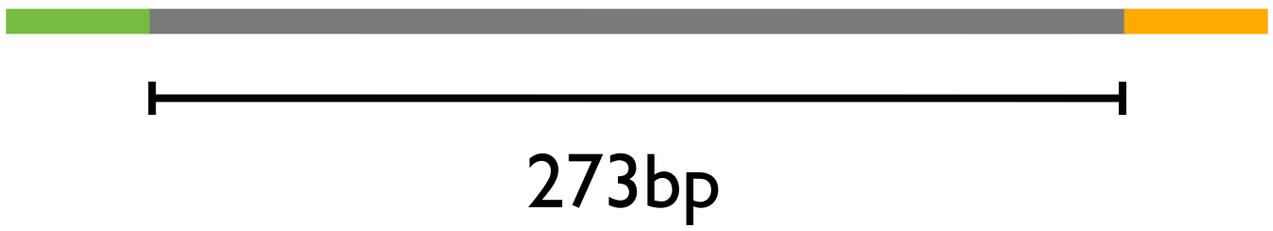
- Universal Priming Sites
  - Sequencing Primers
    - PCR Primers
- Hybridization to Flow Cell
- (more to come)

Additional Sequencing Details

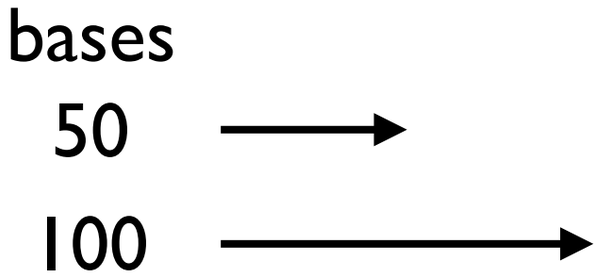
Read Length

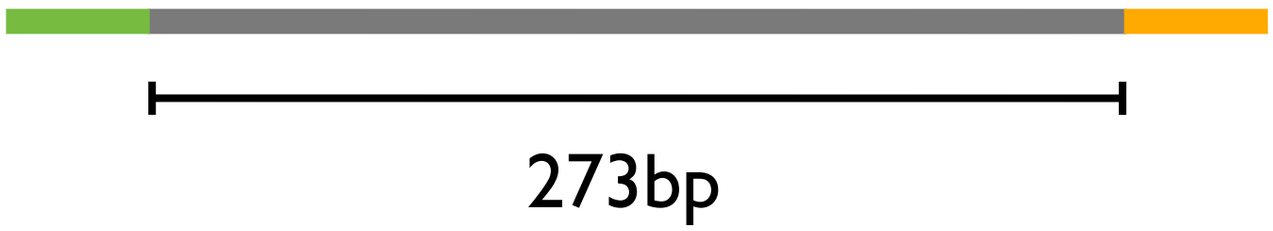
bases

50 ⟶

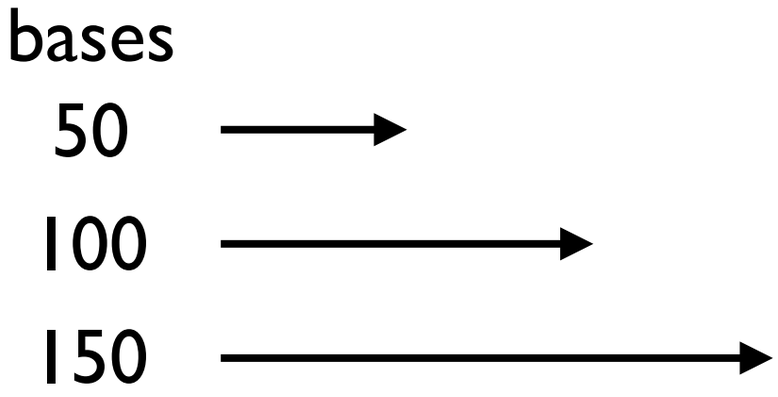


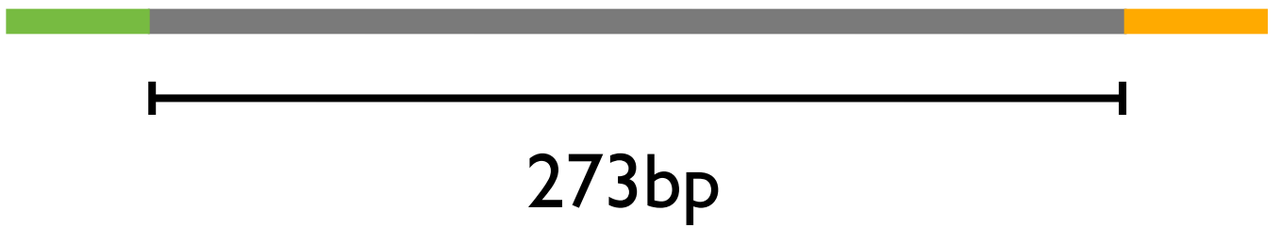
Read Length



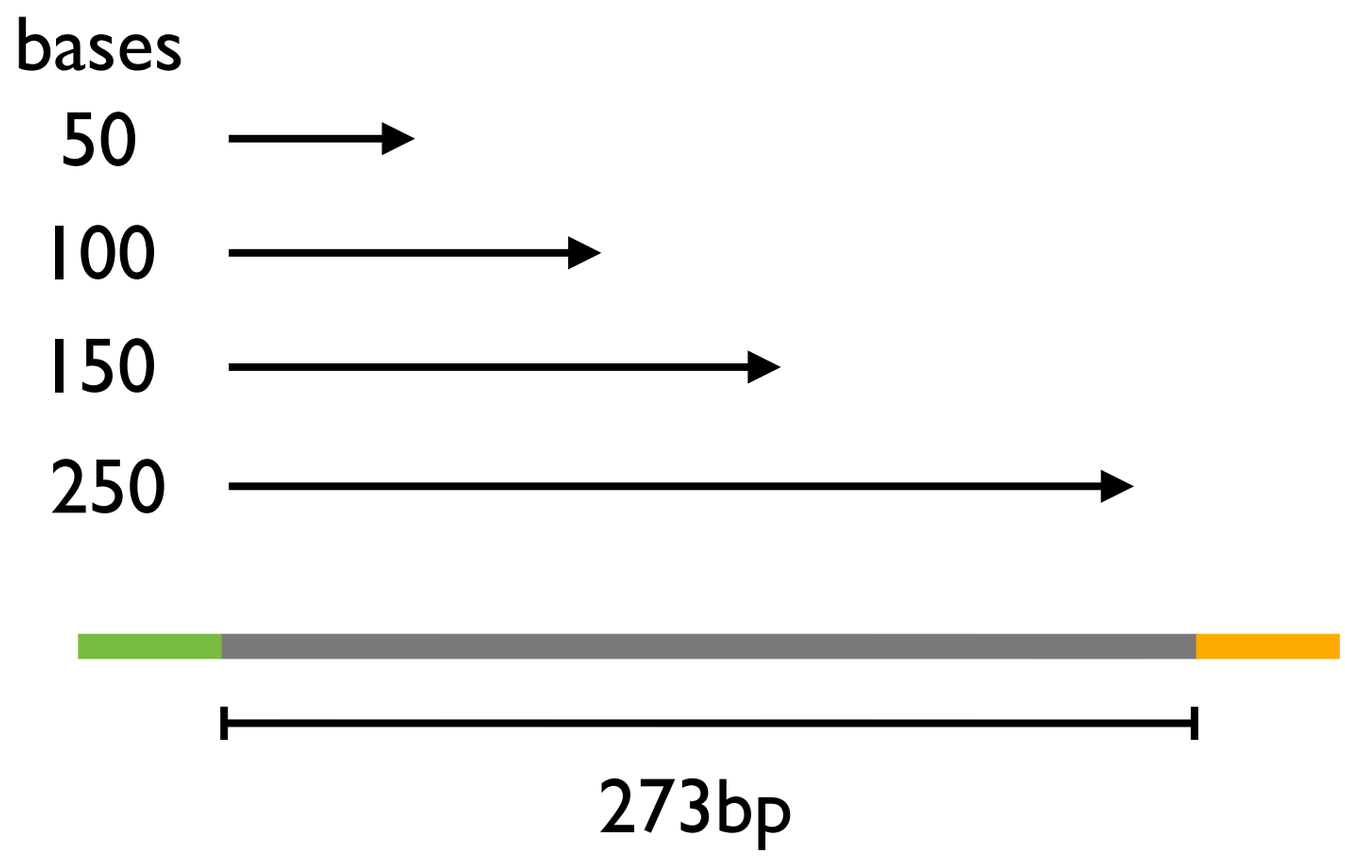


Read Length





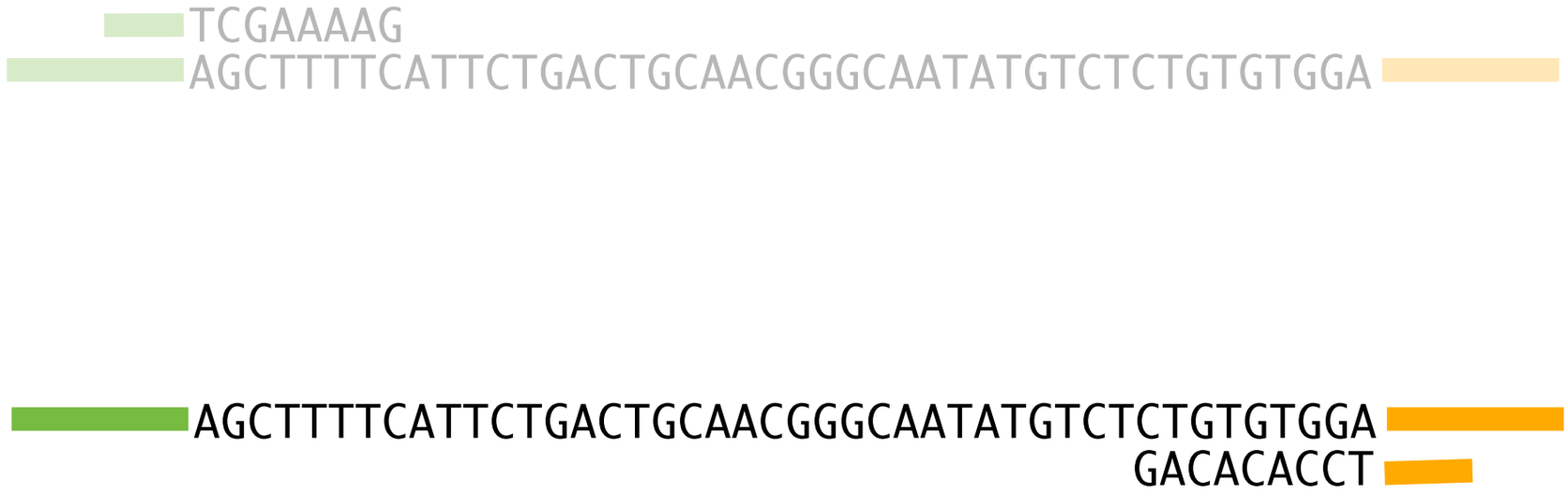
Read Length



Paired-End


TCGAAAAG
AGCTTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGA

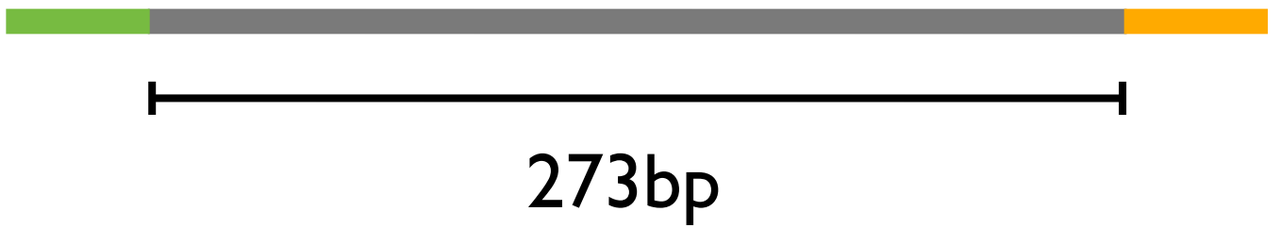
Paired-End


TCGAAAAG
AGCTTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGA
AGCTTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGA
GACACACCT

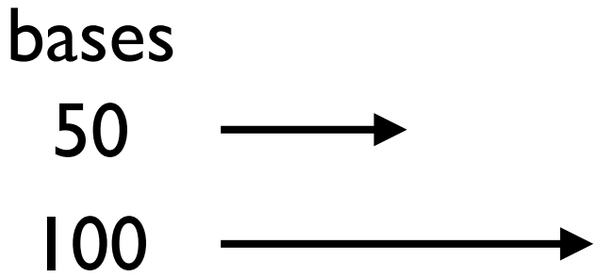
Read Length

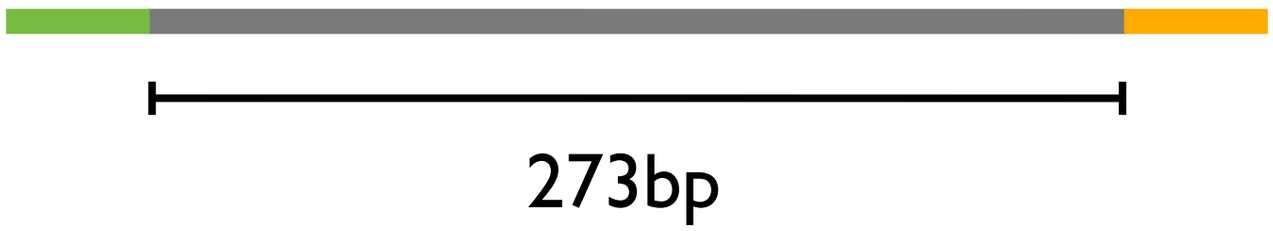
bases

50 ⟶

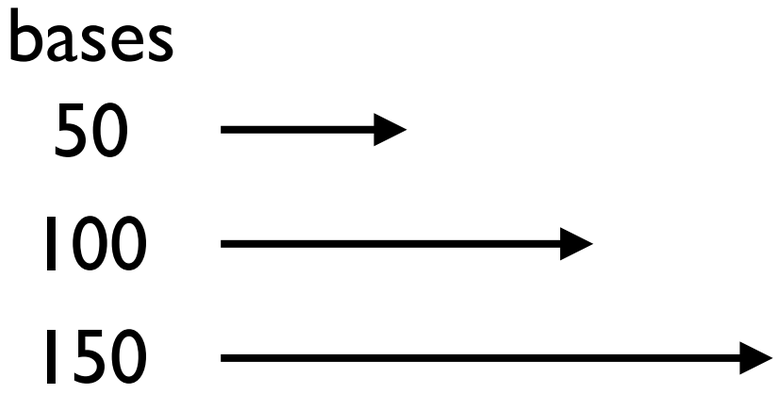


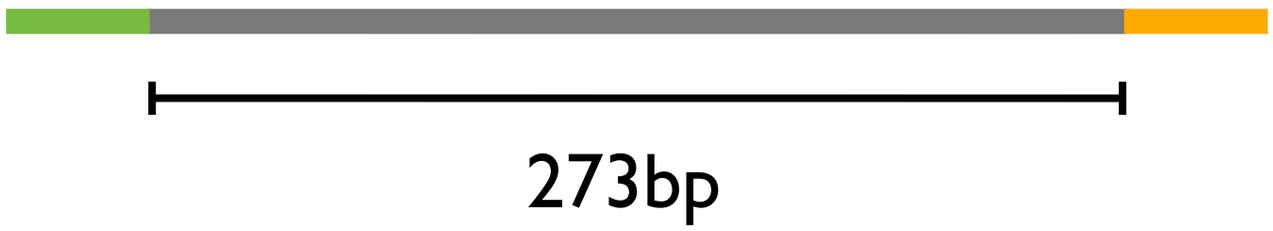
Read Length



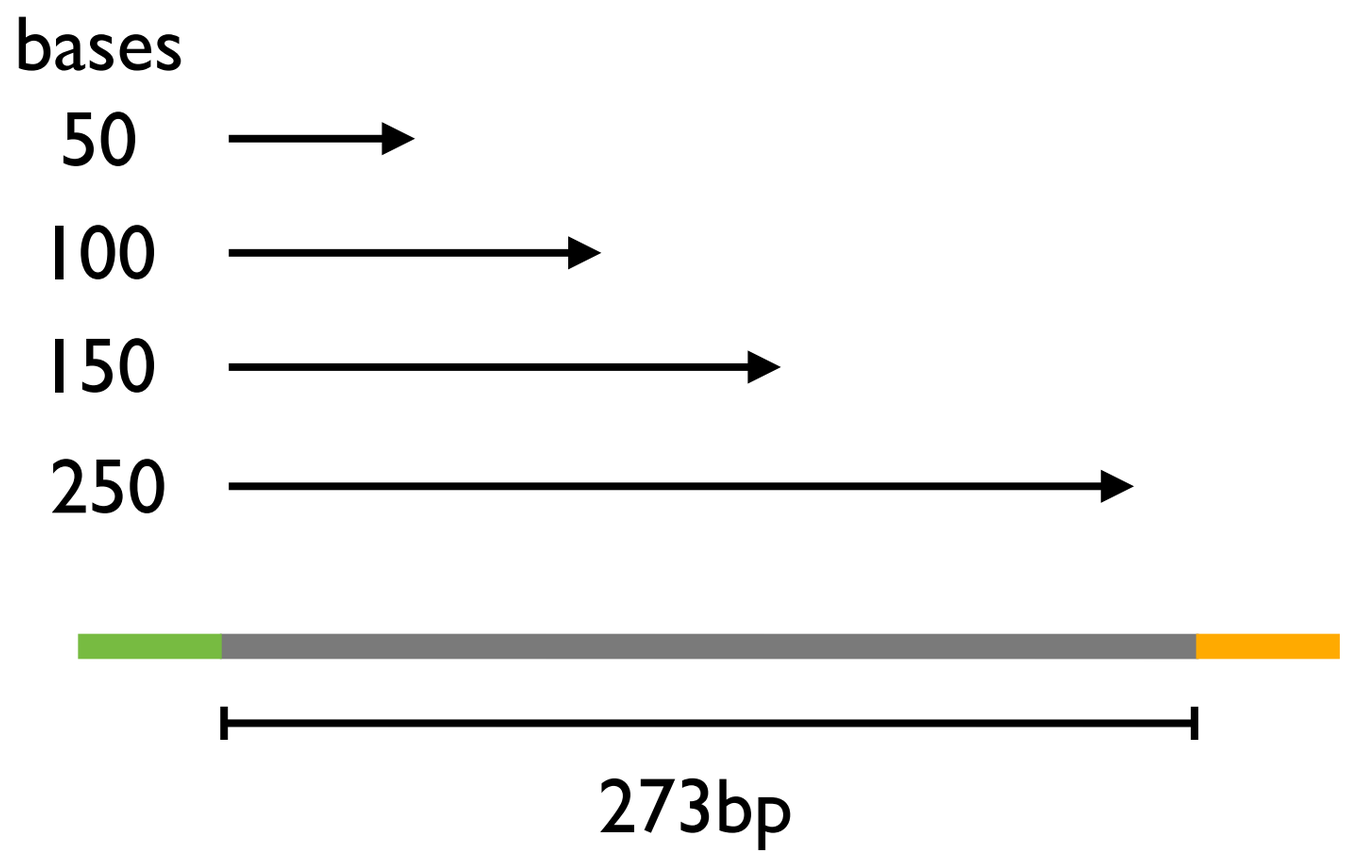


Read Length





Read Length



Read Length

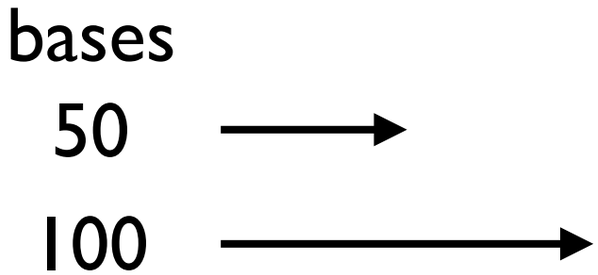
bases

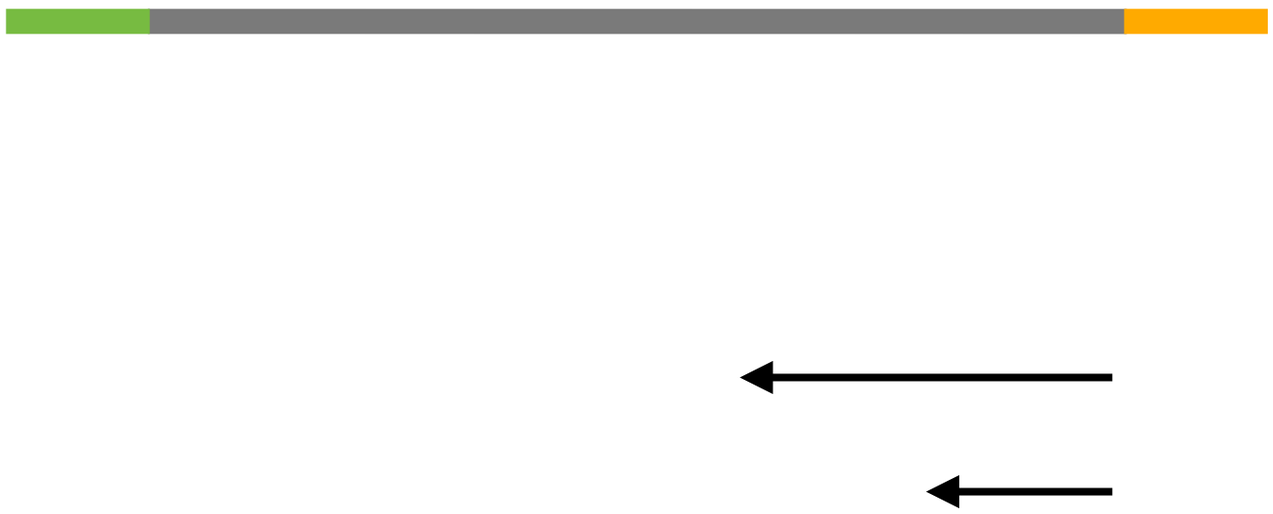
Read Length

bases

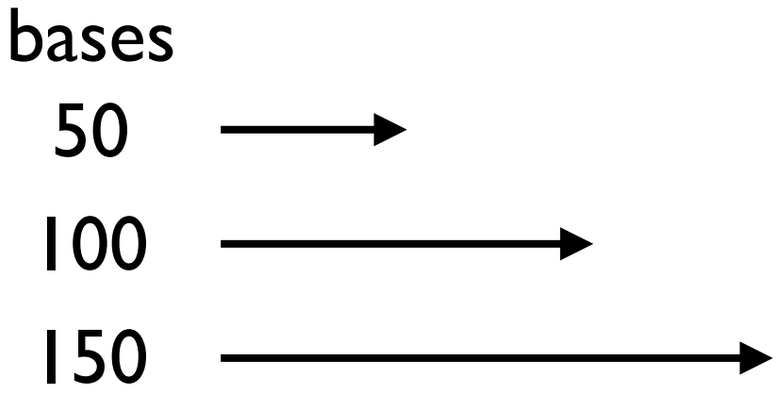
50 ⟶

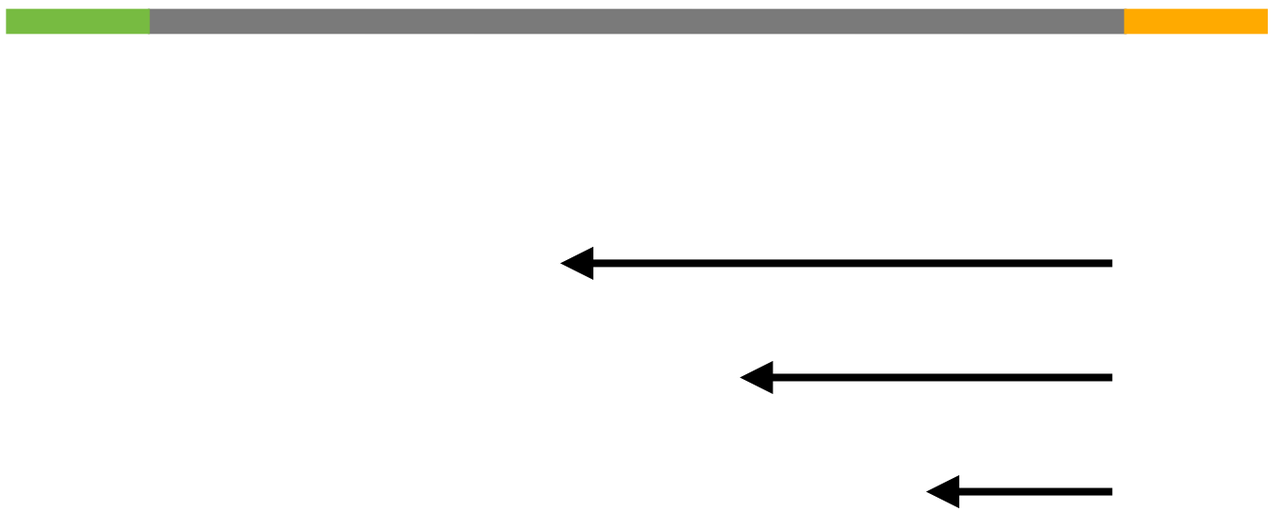
Read Length



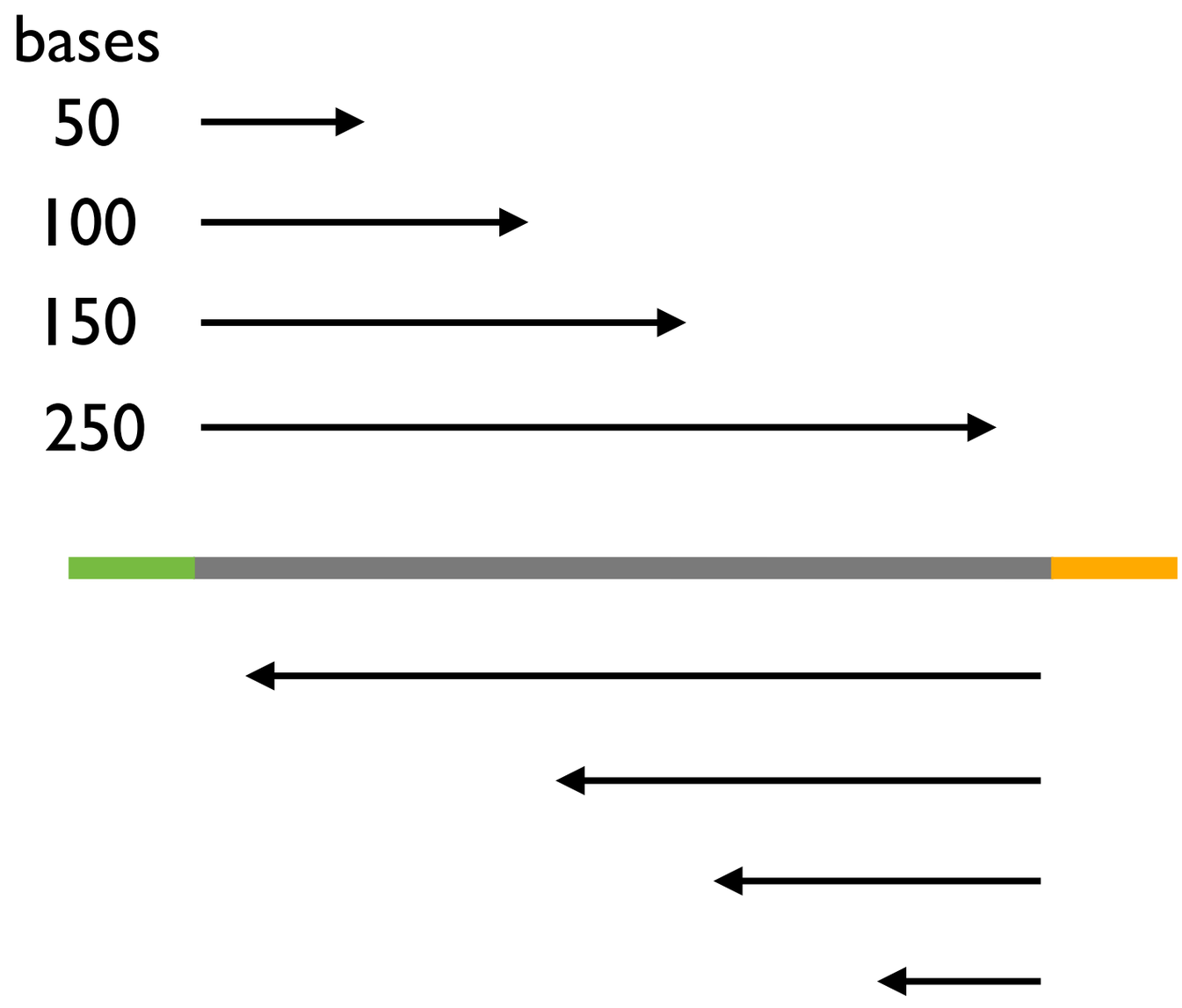


Read Length

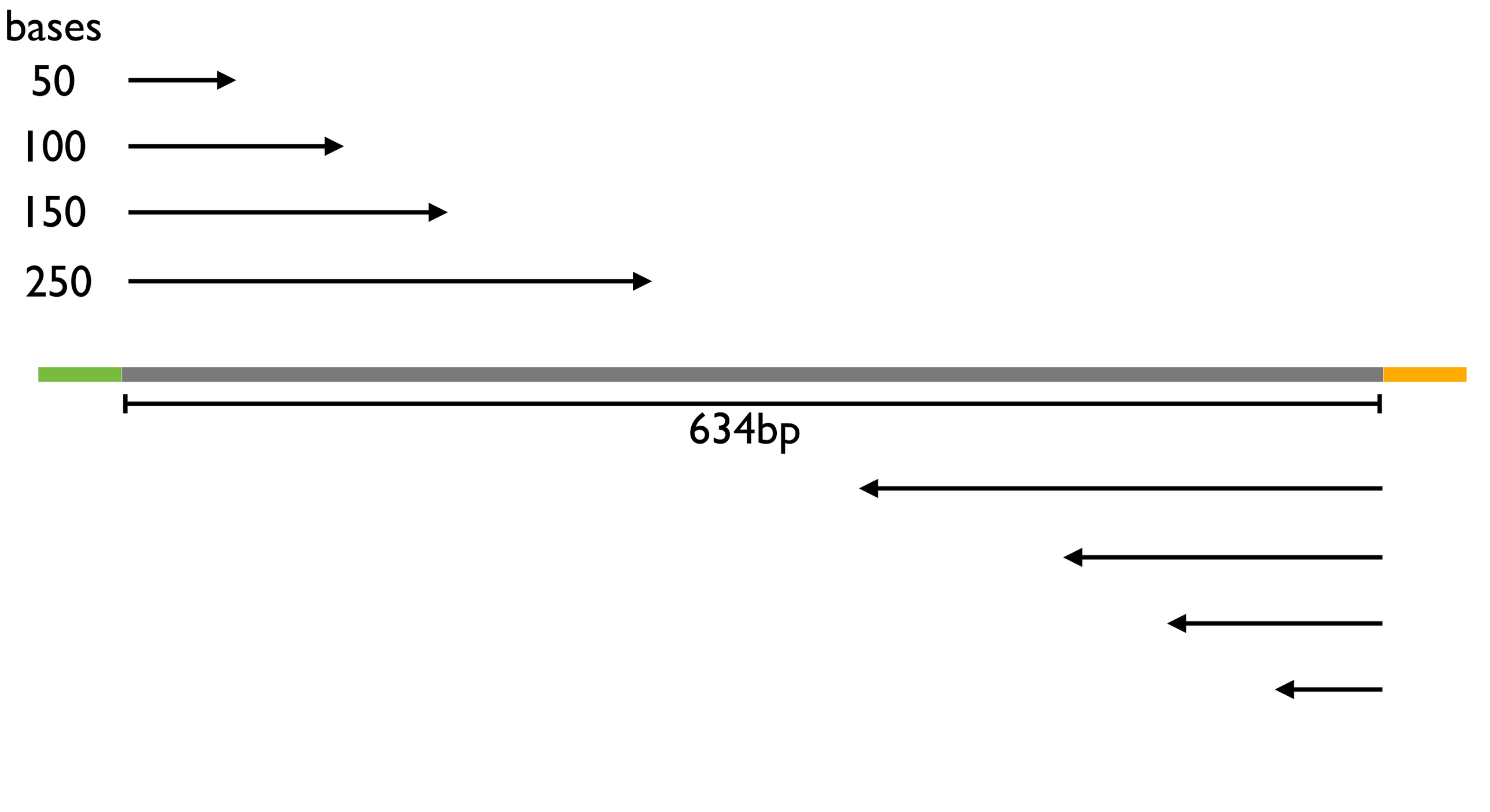




Read Length

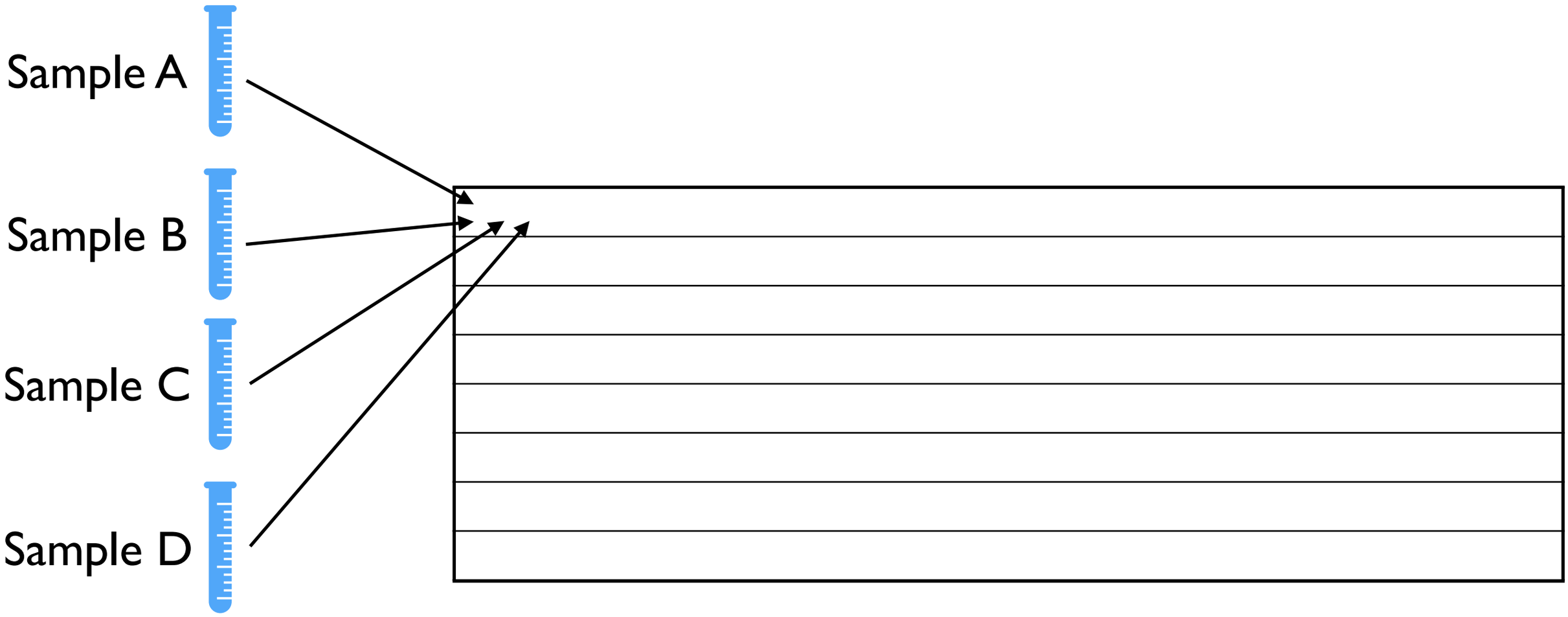


Read Length

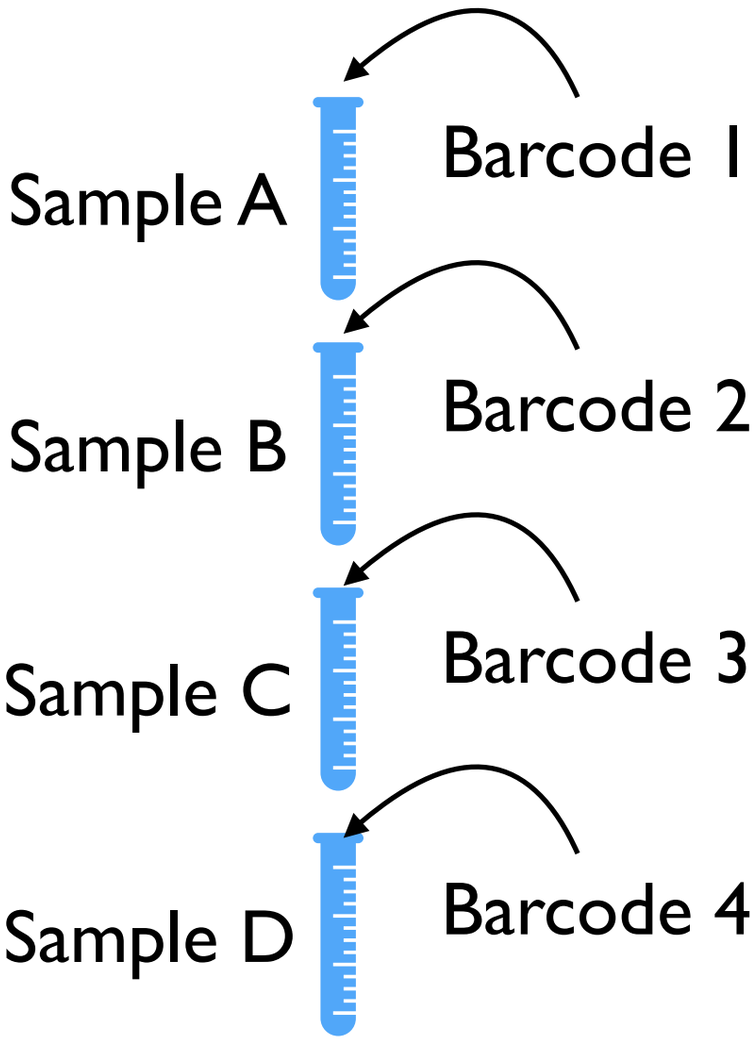


Barcodes: Why?

Multiplexing: Combine multiple samples in a lane



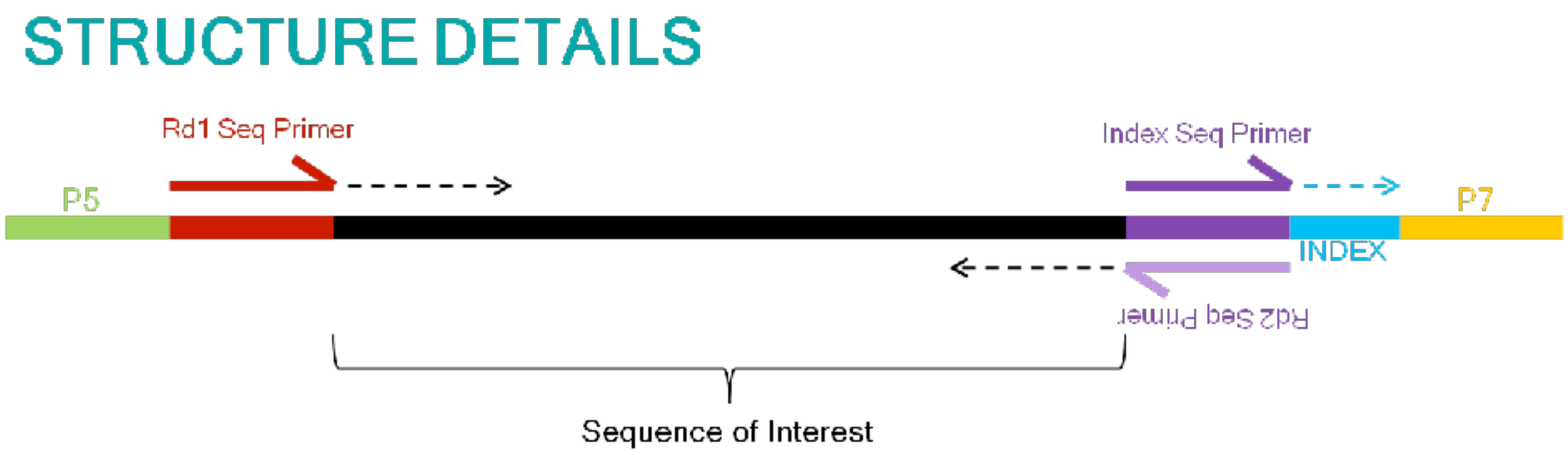
Barcodes


Sample A
Barcode 1
Sample B
Barcode 2
Sample C
Barcode 3
Sample D
Barcode 4

Barcodes

Sample_Name	I7_Index_ID	index
1_A	P49-E1	AAGACCGT
2_A	P50-E2	TTGCGAGA
3_A	P51-E3	GCAATTCC
4_A	P52-E4	GAATCCGT
5_A	P53-E5	CCGCTTAA
6_A	P54-E6	TACCTGCA
7_B	P55-E7	GTCGATTG
8_B	P56-E8	TATGGCAC
9_B	P57-E9	CTCGAACA
10_B	P58-E10	CAACTCCA

Multiplexing (Barcodes)



http://nextgen.mgh.harvard.edu/IlluminaChemistry.html

MiSeq, NextSeq, and More Seqs

	MiSeq	NextSeq	HiSeq 4000	NovaSeq 6000
Maximum Output	15 Gb	120 Gb	750 Gb	3000 Gb
Maximum Reads per Run	25 million	400 million	2.5 billion	10 billion
Maximum Read Length	2 × 300 bp	2 x 150 bp	2 × 150 bp	2 × 150 bp
Run Time	4-56 hours	15-29 hours	< 1–3.5 days	13-45 hours
Cost*	$1,787	$4,695	$19,206	$35,538
Cost/Mbp*	$0.119	$0.039	$0.026	$0.012

* Duke Sequencing and Genomic Technologies Shared Resource, July 2018

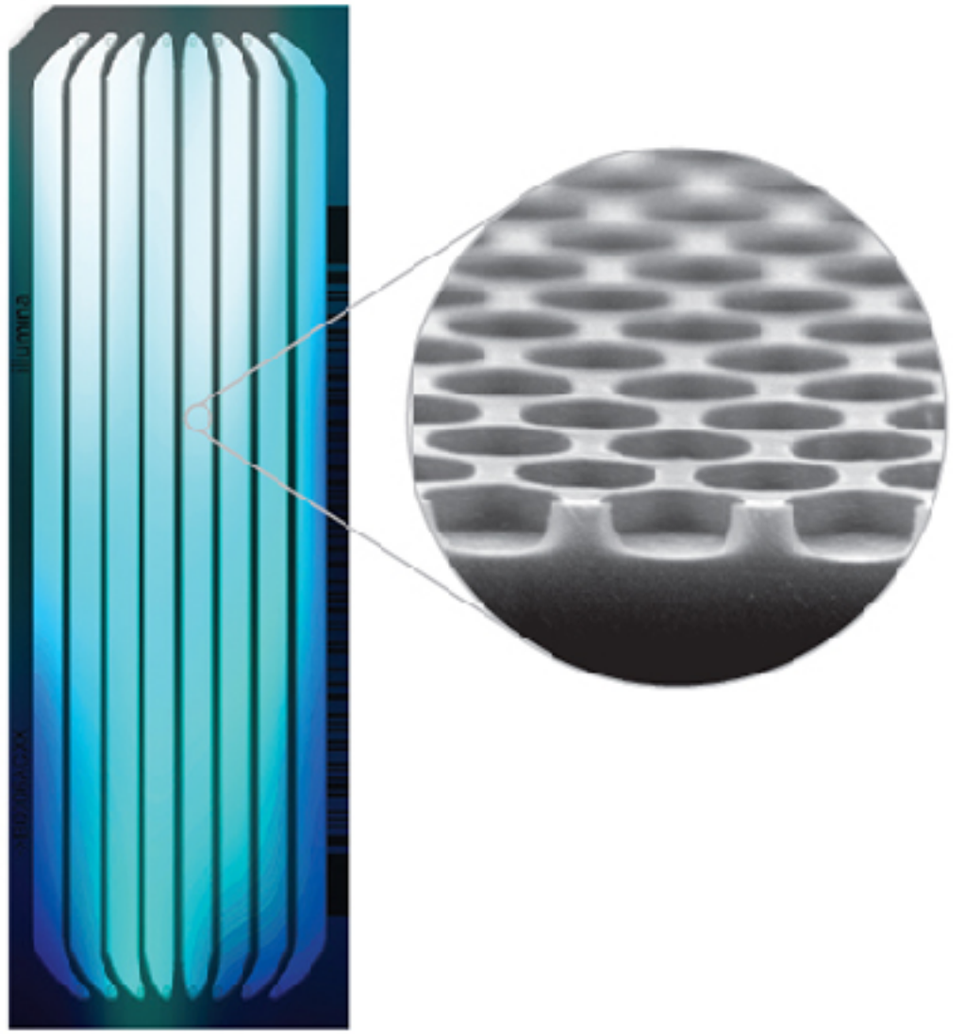
1st Generation	2nd Generation	3rd Generation
Chemical (Maxim-Gilbert)	Pyrosequencing (454)	Single molecule real time (PacBio)
Chain Termination (Sanger)	Chain Termination (Illumina)	Nanopore sequencing (Oxford Nanopore)
Pyrosequencing	Sequencing by ligation (SOLiD sequencing)	
	Ion semiconductor (Ion Torrent)	

Illumina Video

https://www.youtube.com/watch?v=HMyCqWhwB8E

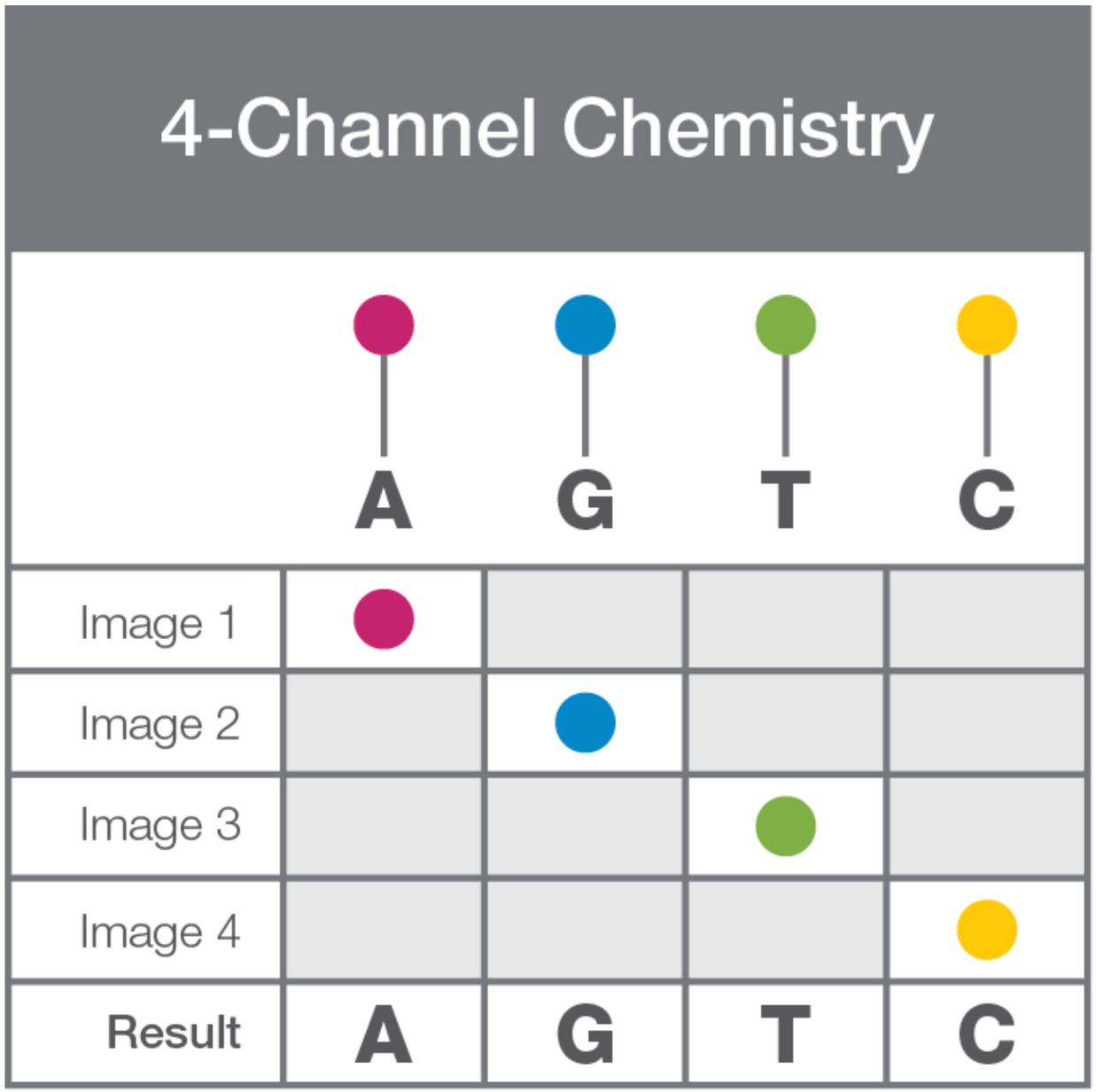
Patterned Flow Cells

- ExAmp
- Machines
  - HiSeq X
  - HiSeq 3000/4000
  - NovaSeq 6000



https://assets.illumina.com/content/dam/illumina-marketing/images/science/v2/web-graphic/nano-well-technology-web-graphic.jpg

4-Channel Chemistry



https://www.illumina.com/content/dam/illumina-marketing/documents/products/techspotlights/cmos-tech-note-770-2013-054.pdf

2-Channel Chemistry

4-Channel Chemistry

	A	G	T	C
Image 1	●			
Image 2		●		
Image 3			●	
Image 4				●
Result	**A**	**G**	**T**	**C**

2-Channel Chemistry

	A	G	T	C
Image 1	●		●	
Image 2	●			●
Result	**A**	**G**	**T**	**C**

https://www.illumina.com/content/dam/illumina-marketing/documents/products/techspotlights/cmos-tech-note-770-2013-054.pdf