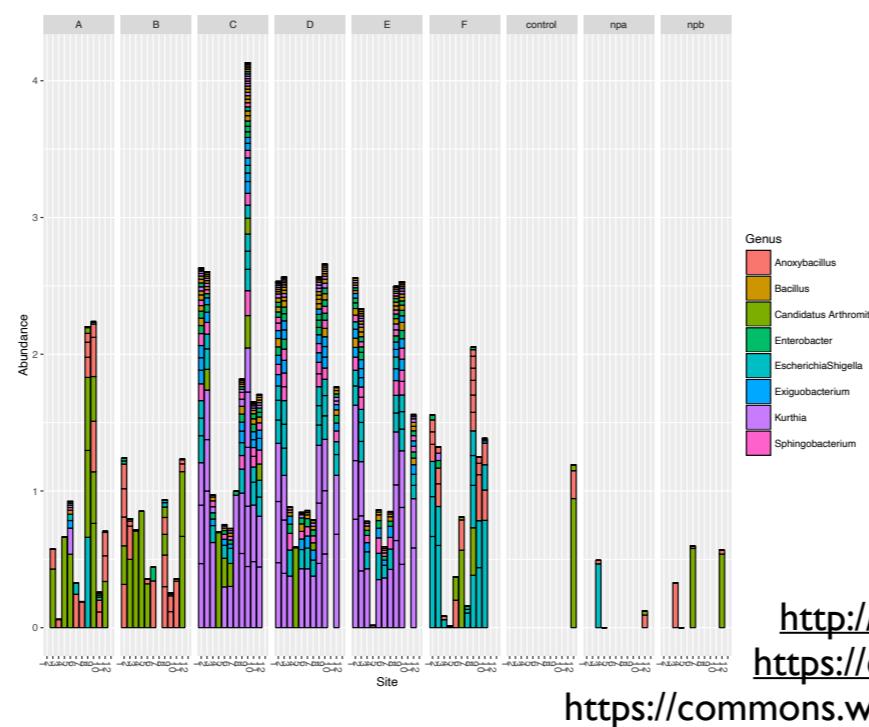


HTS Background and Theory

Josh Granek

Molecular Biology

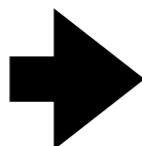


<http://www.geograph.org.uk/photo/2847164>

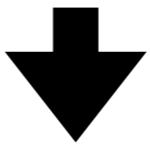
<https://commons.wikimedia.org/wiki/File:Pcr.jpg>

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

Bioinformatic Analysis



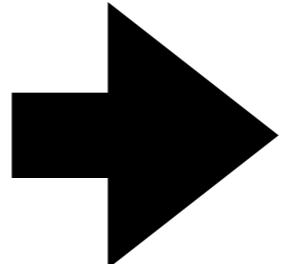
```
@M00698:36:000000000-AFBEL:1:1101:14738:1412 1:N:0:0
TTACGCTAACAGGCAGGTAGCCTGGCAGGGTCAGGAAATCAATTAACTCATCGGAAGTGGTATCTGTTCCATCAAGCGTGCAGCTCGTCAAAACGCC
+
ABBBABBBBAFFGGGGGGGGGGHGGHGGCG2GF3FFGHHHHHGGFGHEHHGGGEHHHAGGHHGHHFFDHFHHGECCCC@F@H?GHH/GBEFGGG
@M00698:36:000000000-AFBEL:1:1101:16483:1412 1:N:0:0
CTGCCAGTTGAACGACGGCGAGCAGTTAAAGCCAGCAGTTGCCGGATATTCGCGTGGATAGCTTGCAAAGCGACGCCAGTTCCAGATCCGGCG
+
AAABBFFFFFFGGGGGGGGGGGGHHHHHHHHGHGHGHGHGHGHGGGGHHHHGGGGGGHHHHFFHHHHGHGGGGGGGGHHHHHHGGGGGGGGGGGGGGGG
@M00698:36:000000000-AFBEL:1:1101:15928:1413 1:N:0:0
GTAAAGTCTGAGTGATACCGCAACTTTACCCCCAGTCCCACTTCGAACCCGAAACATATCGGAAAGAGGCCGTGCCTGATTTAAAGCCGTAGGT
+
```



	Sample 1	Sample 2	...	Sample N
Bacteria 1				
Bacteria 2				
...				
Bacteria N				

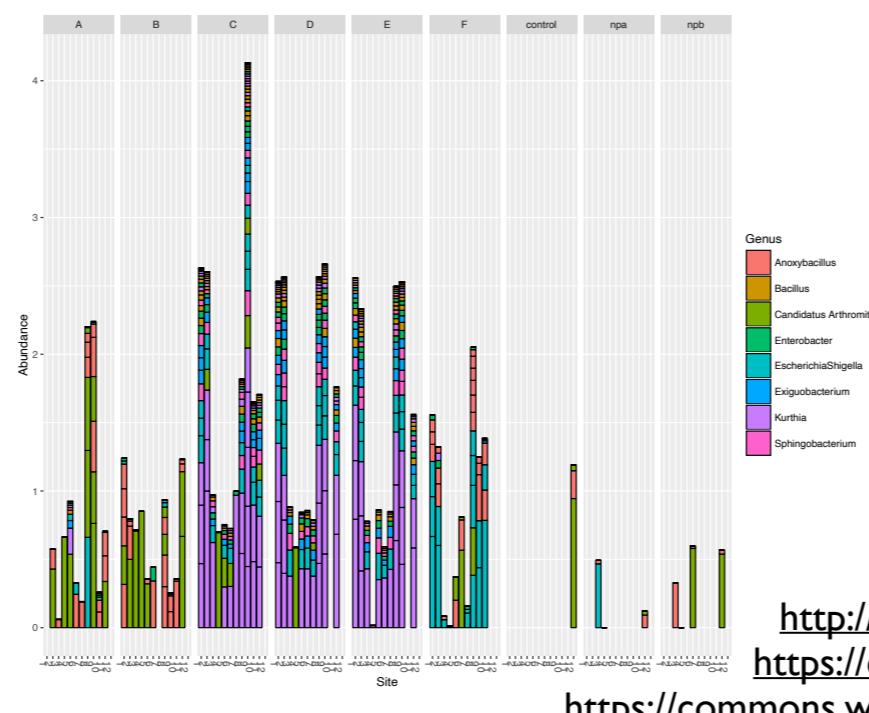
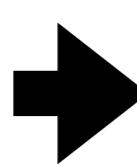
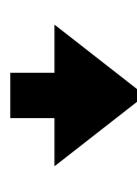
Statistical Analysis

	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

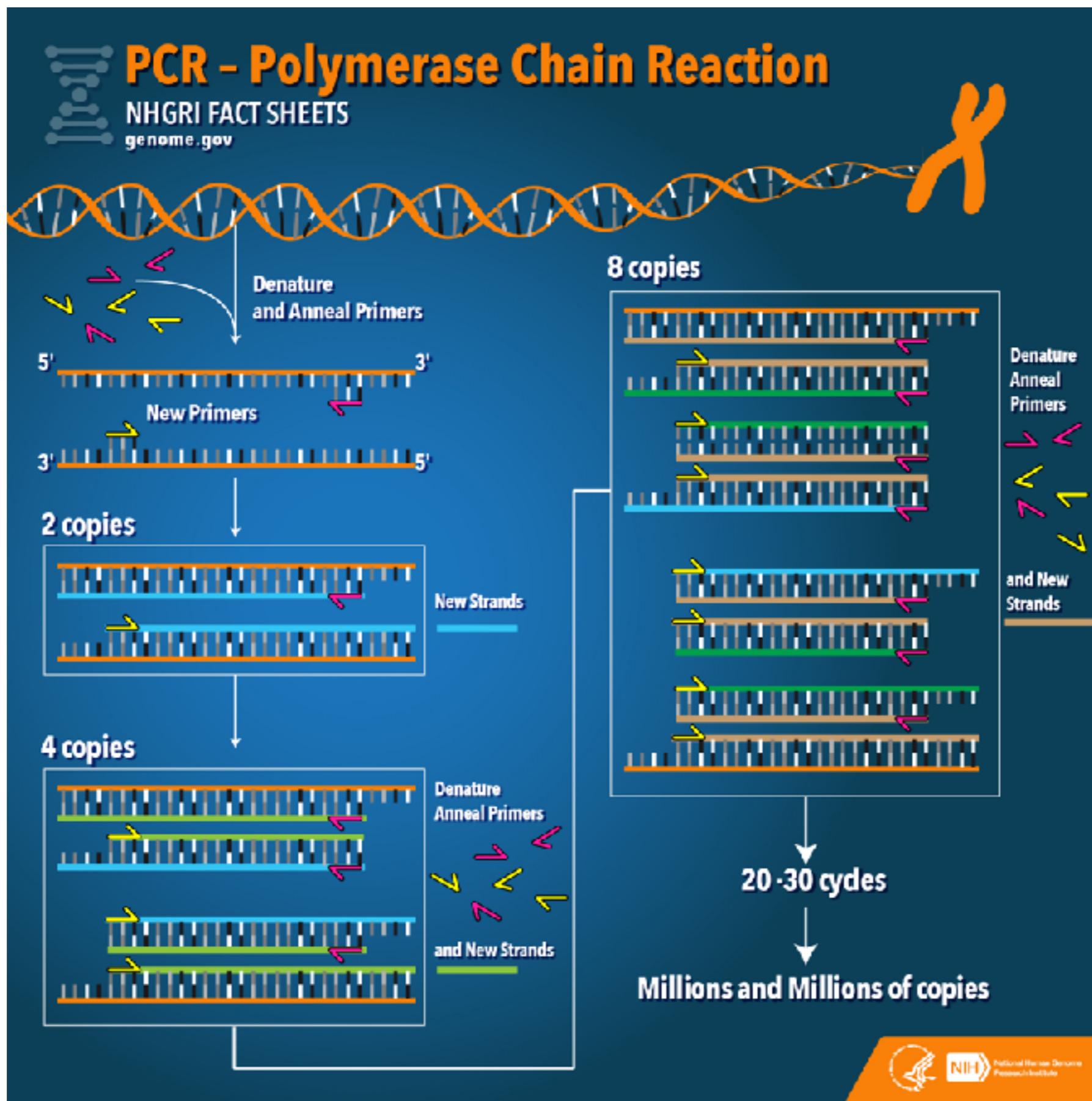


<http://www.geograph.org.uk/photo/2847164>

<https://commons.wikimedia.org/wiki/File:Pcr.jpg>

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

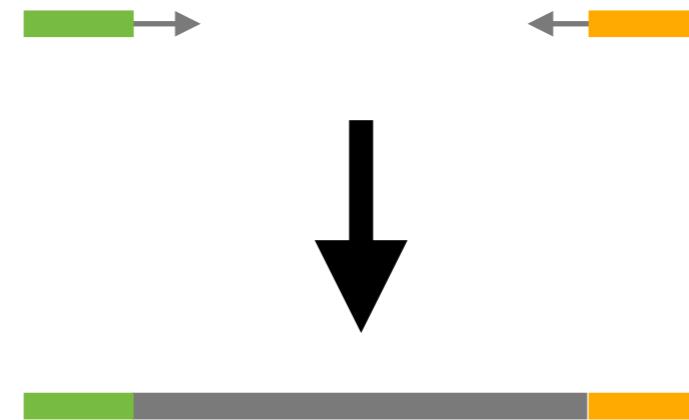
PCR



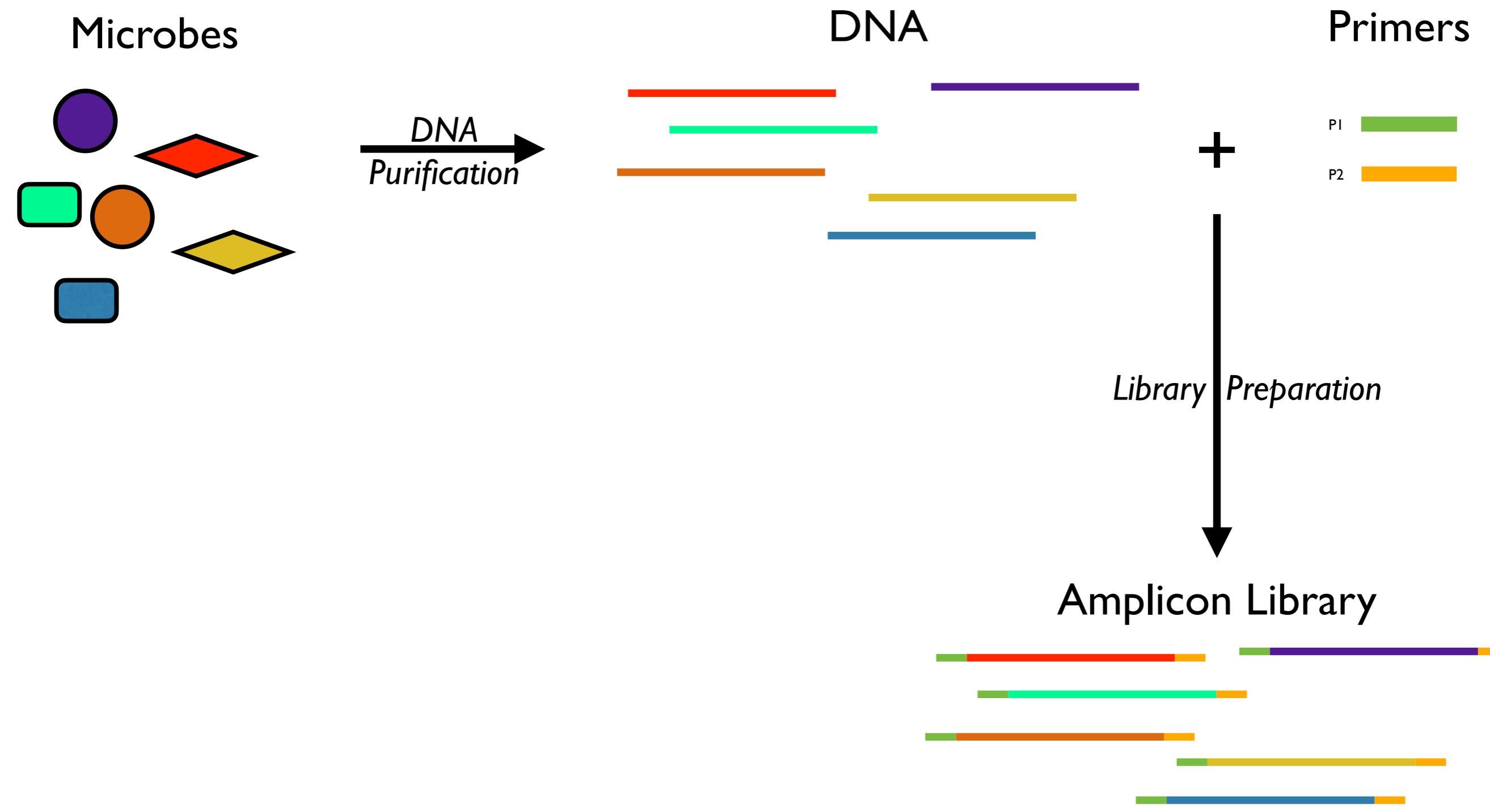
Purified DNA



PCR Amplification



Molecular Biology

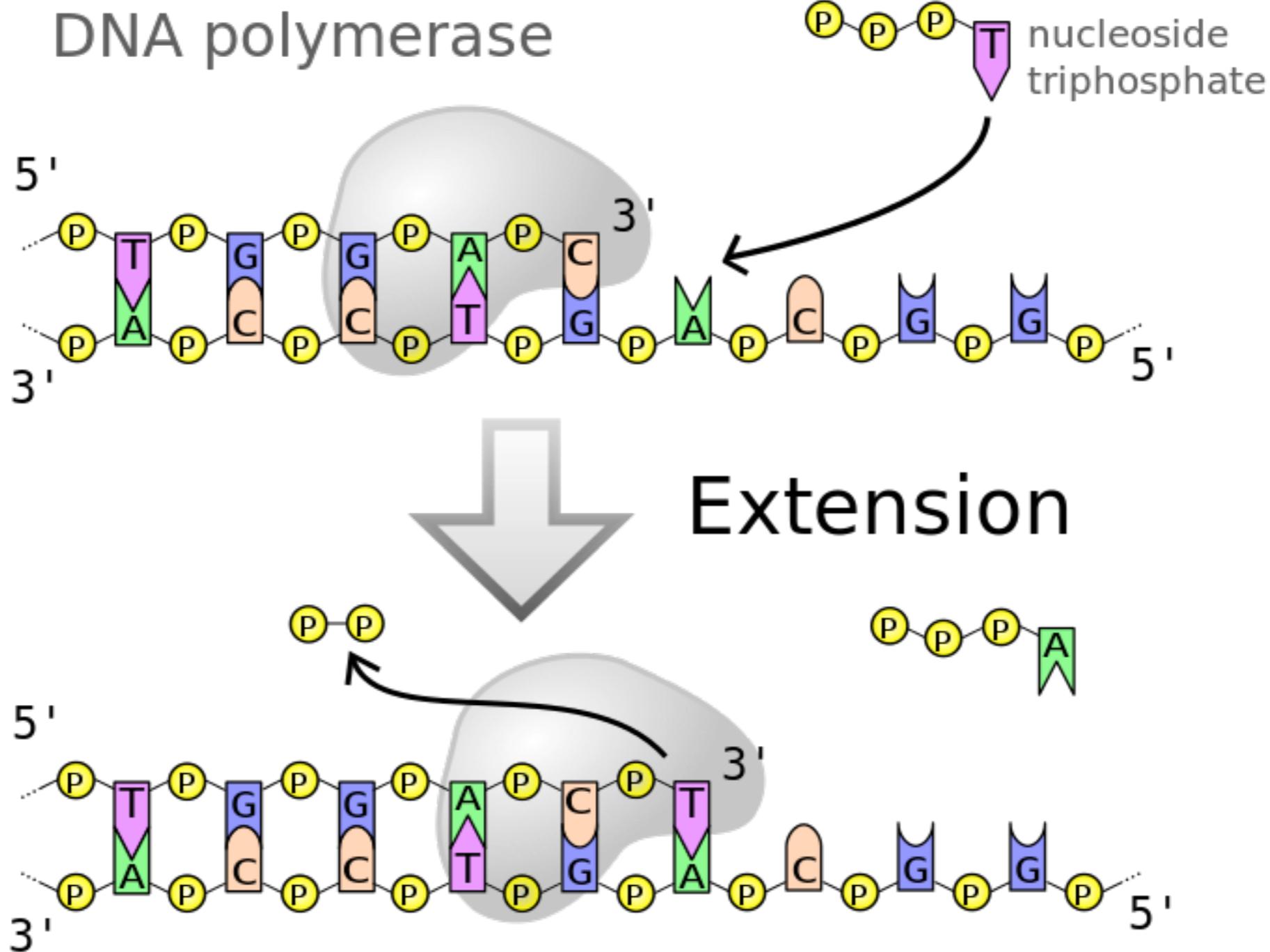


Sanger Sequencing

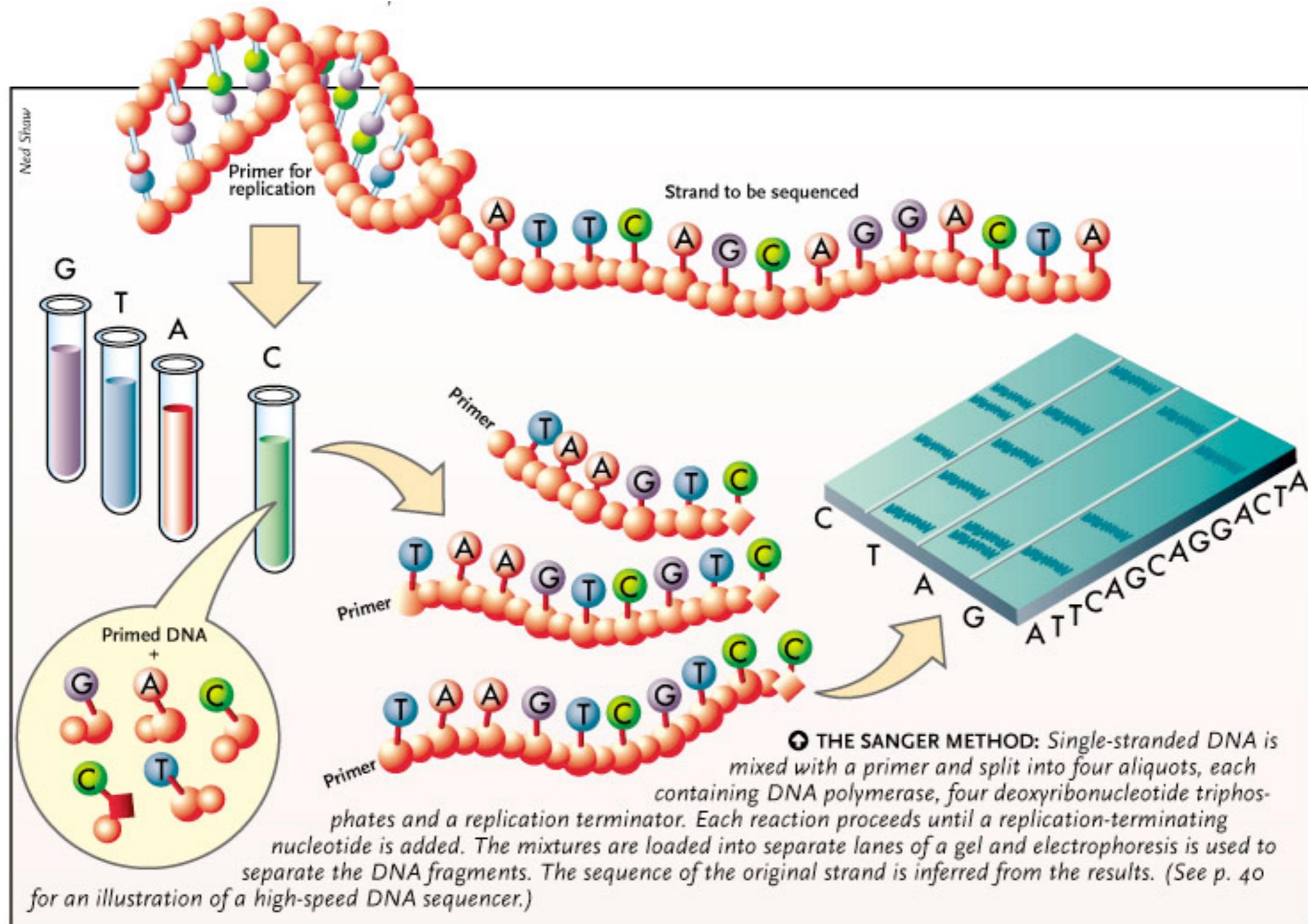
DNA Synthesis

- What are the minimum components for DNA Replication?

DNA Synthesis



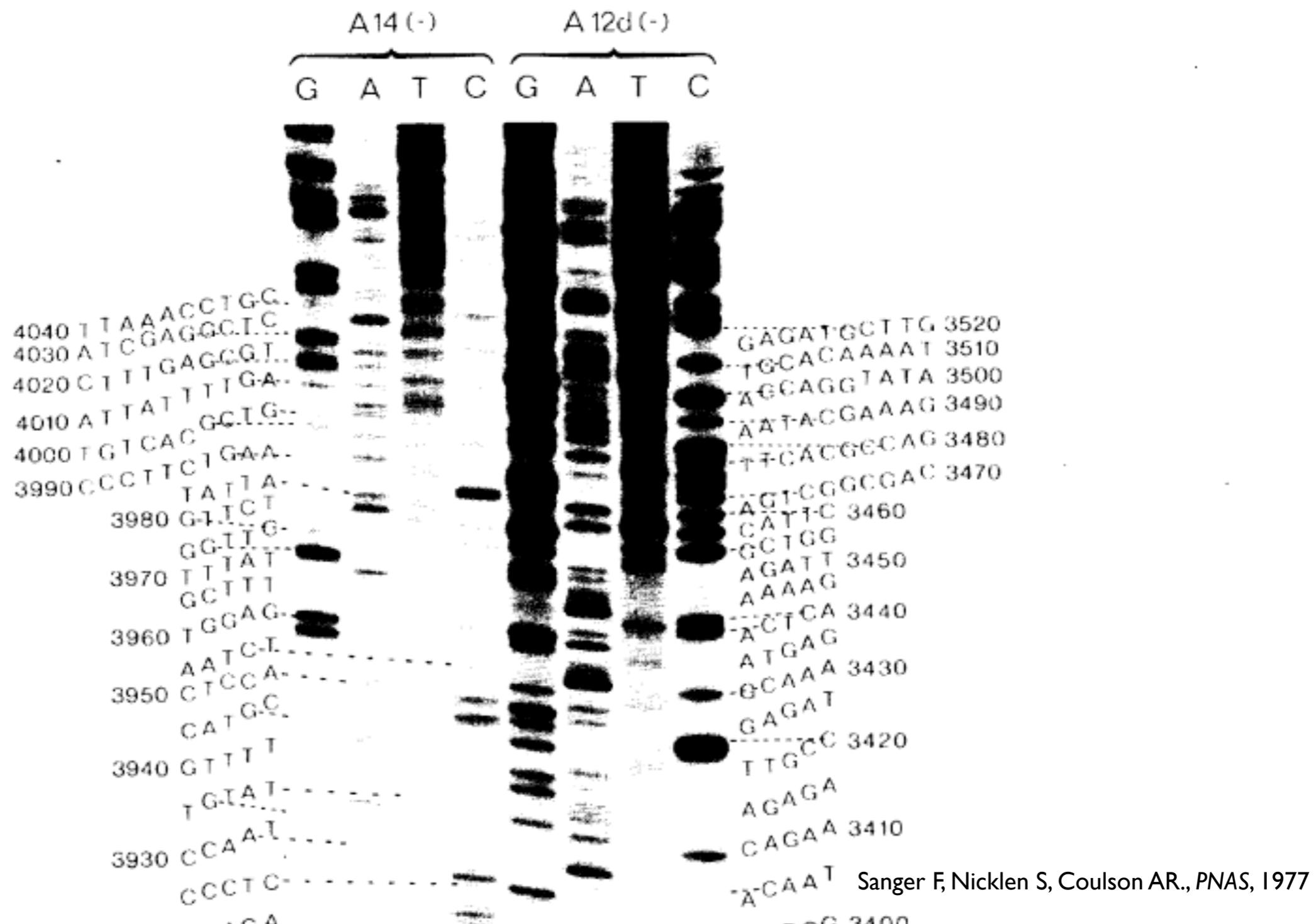
Sanger Sequencing



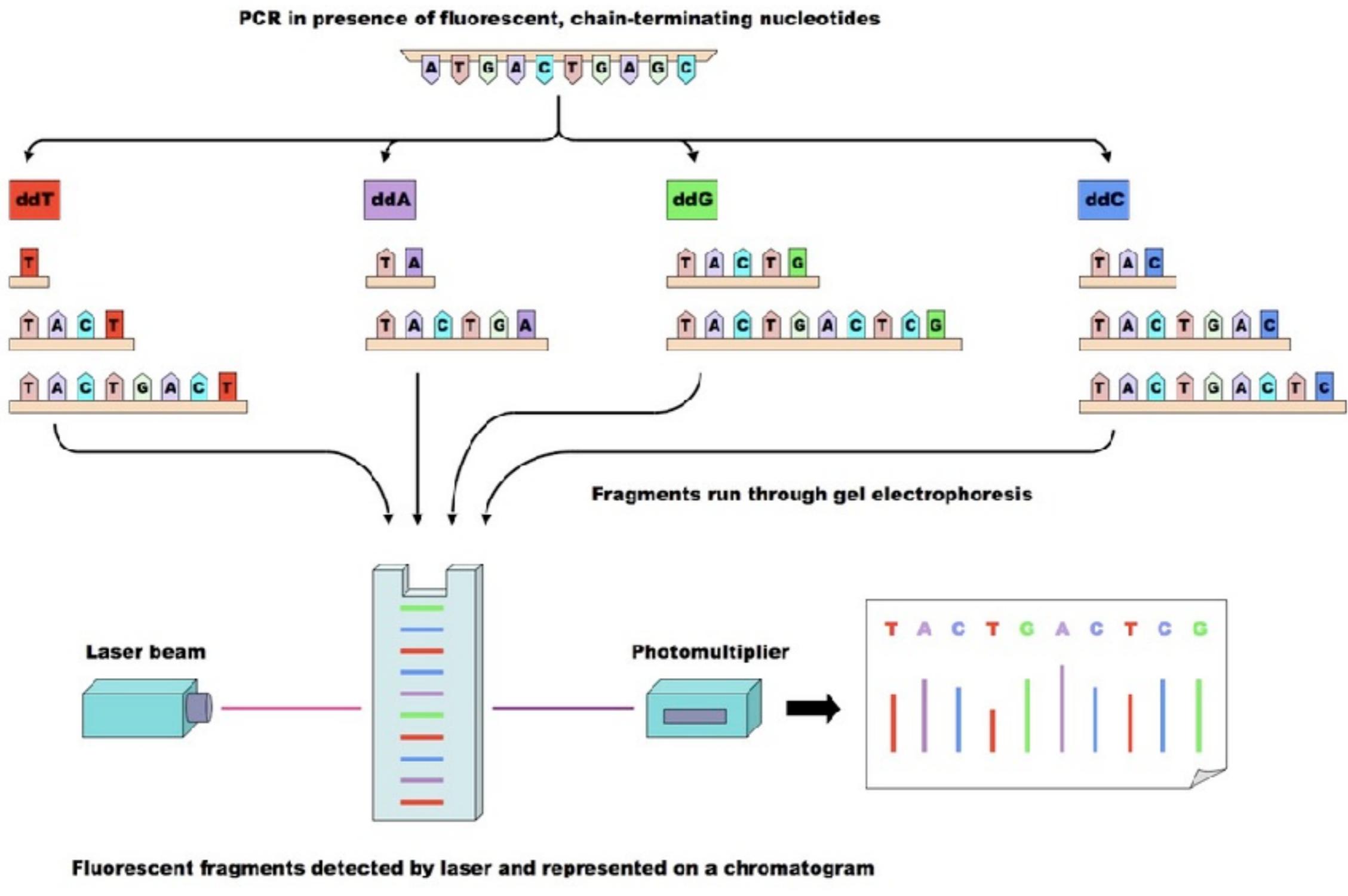
Sanger Sequencing

5464 Biochemistry: Sanger *et al.*

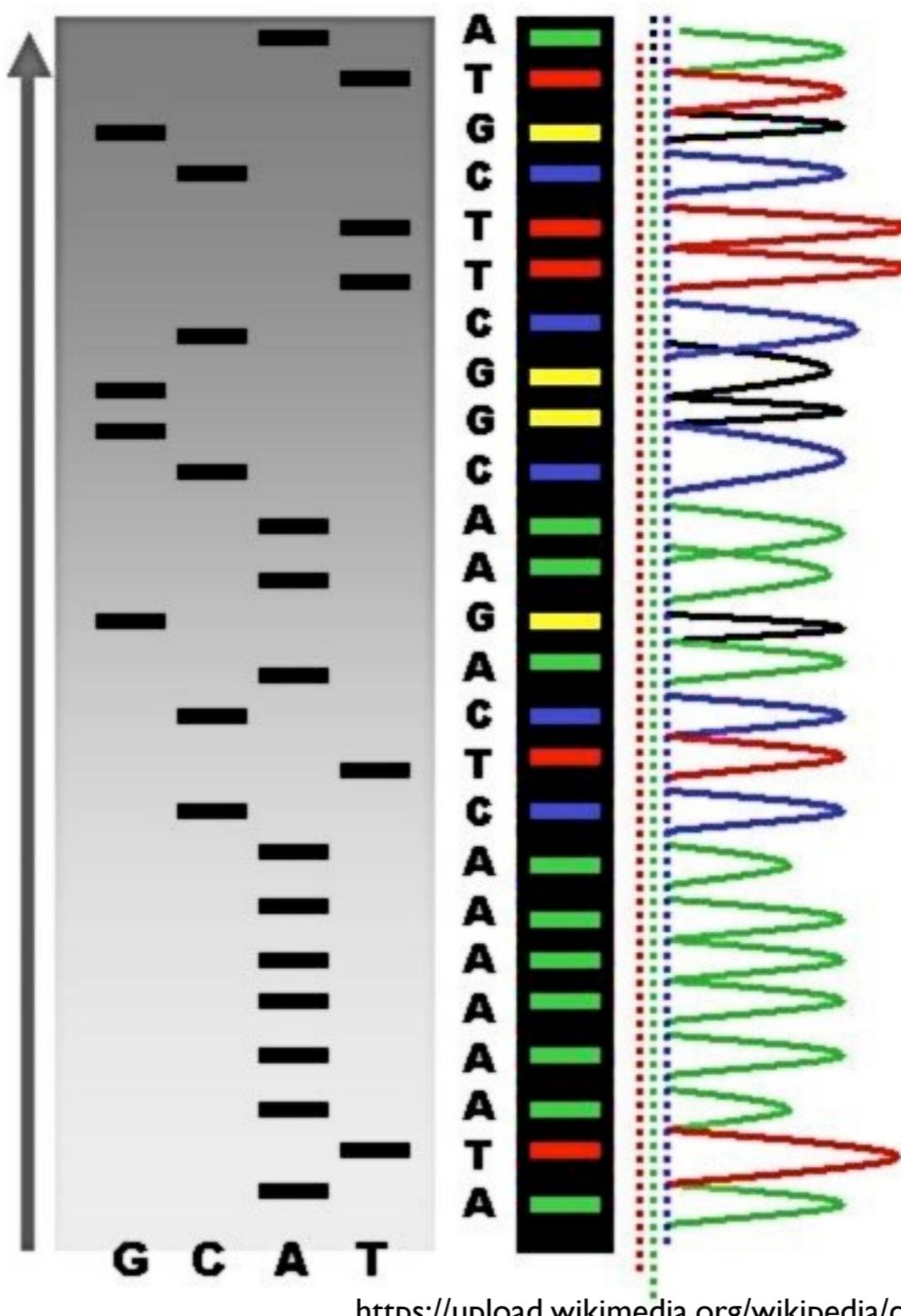
Proc. Natl. Acad. Sci. USA 74 (1977)



Dye-terminator

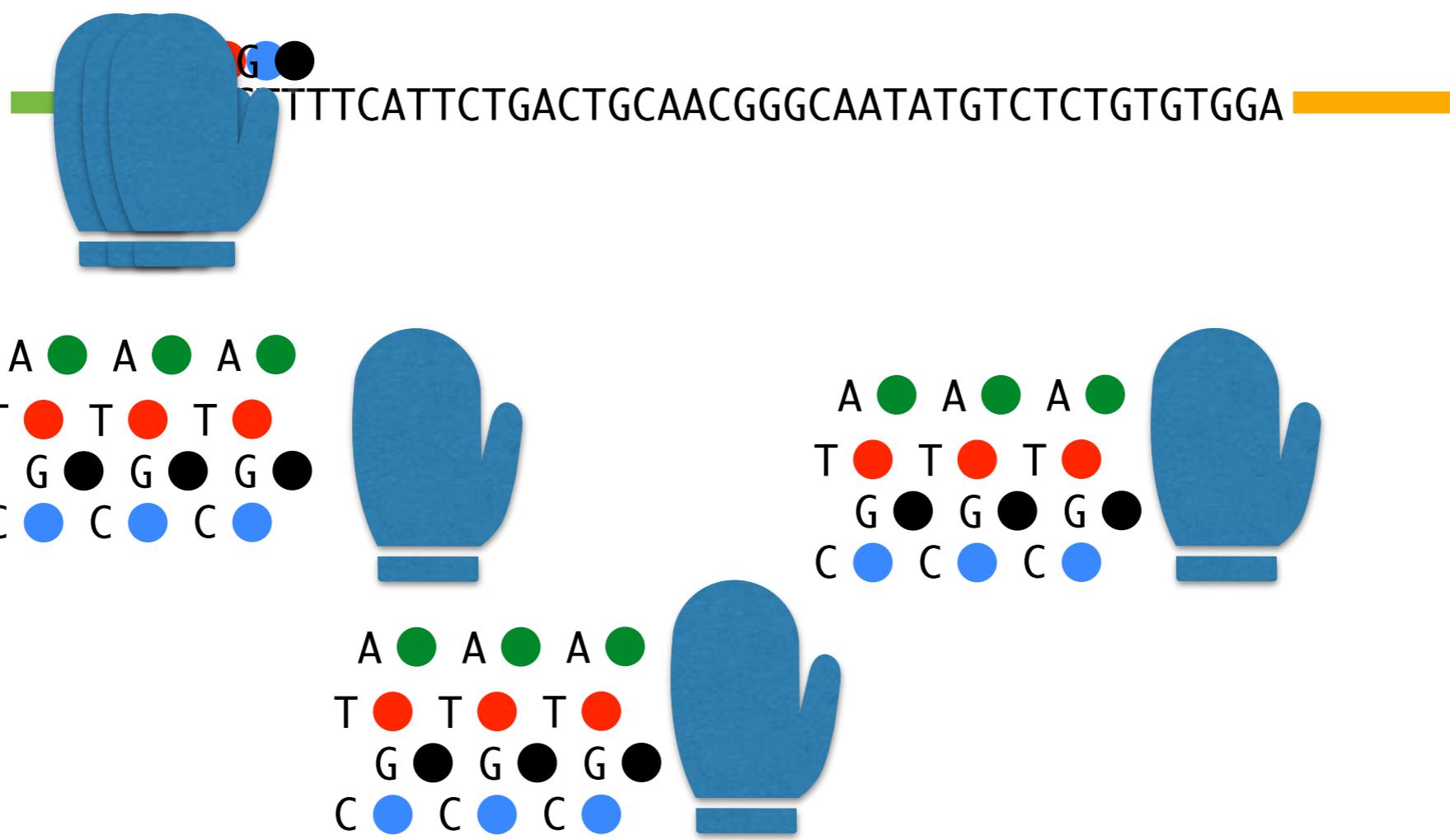


Radiolabel vs. Dye

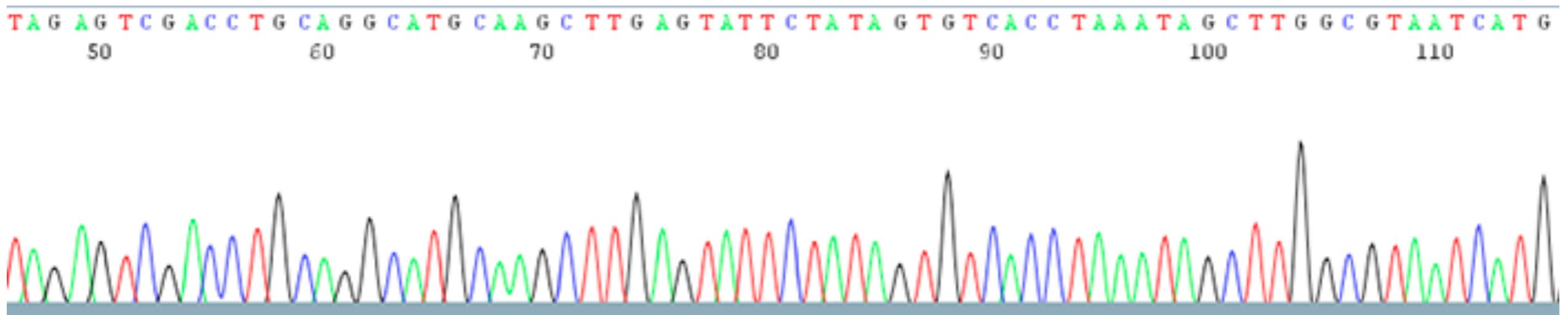


High-Throughput Sequencing

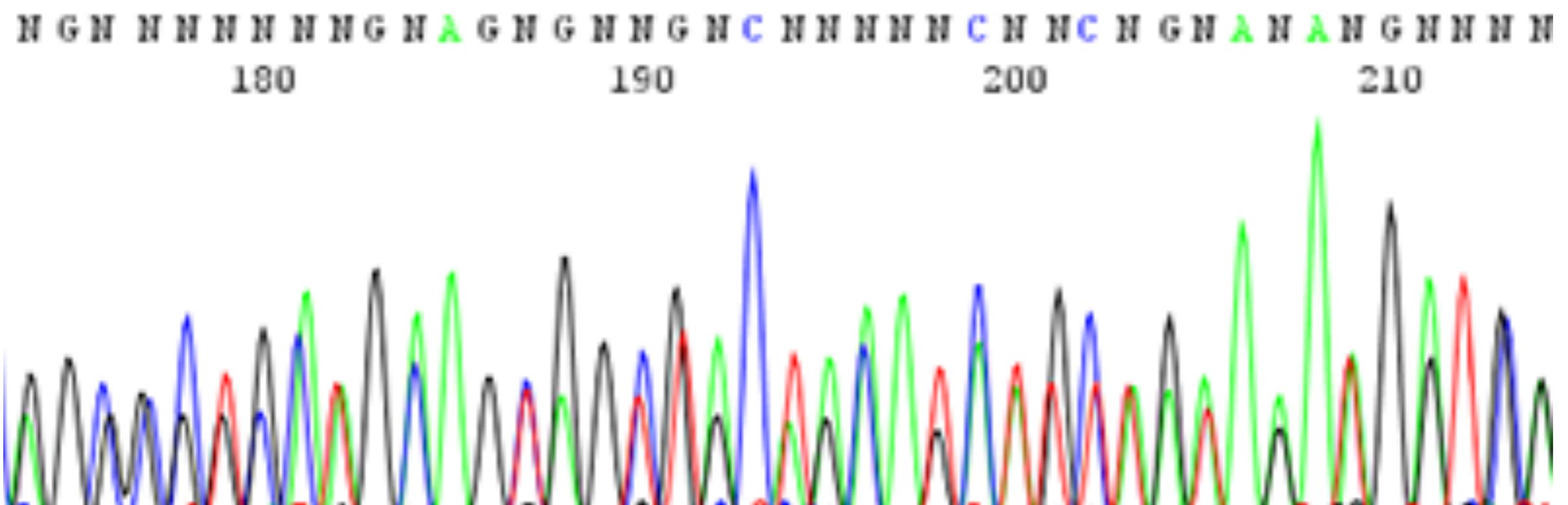
Sequencing



Dye-terminator Sanger Sequencing



Double Sequence



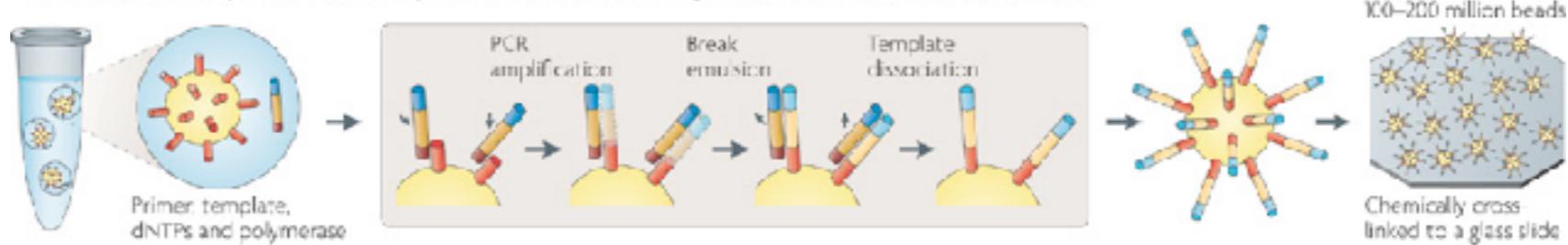
How?

- Separate
- Detect
- Removable Terminator

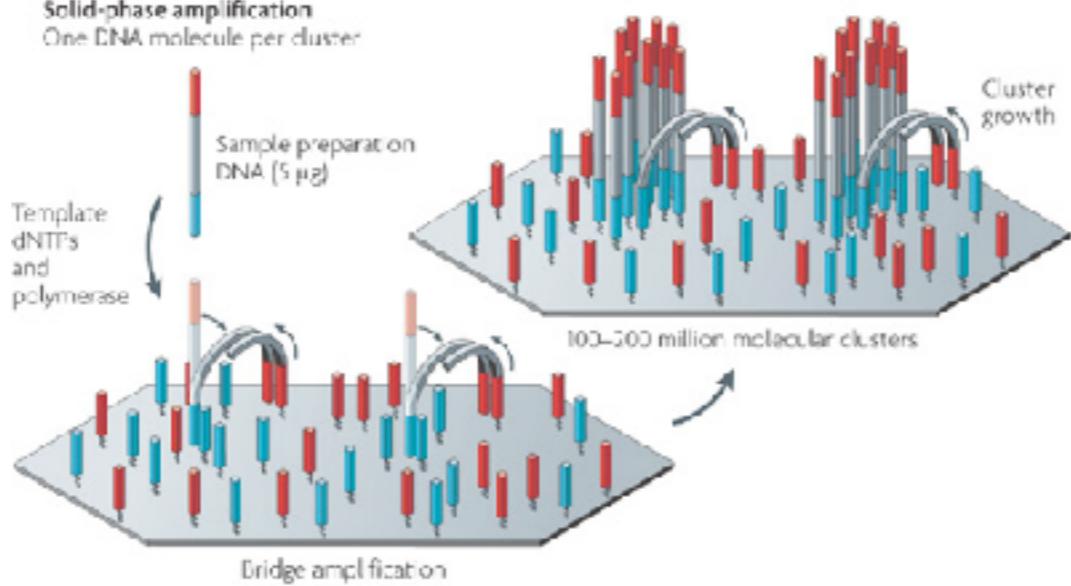
Template immobilization

a Roche/454, Life/APG, Polonator
Emulsion PCR

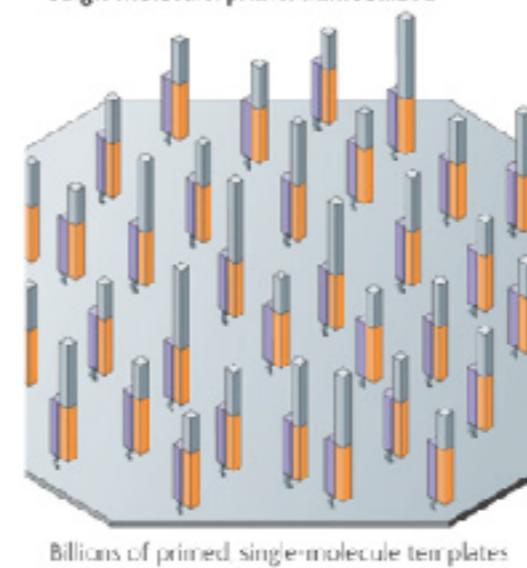
One DNA molecule per bead. Clonal amplification to thousands of copies occurs in microreactors in an emulsion



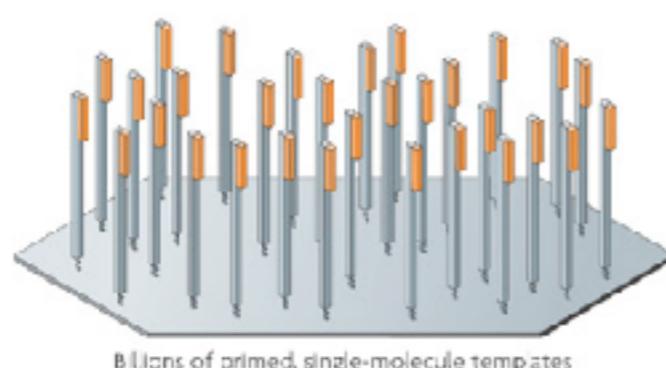
b Illumina/Solexa
Solid-phase amplification
One DNA molecule per cluster



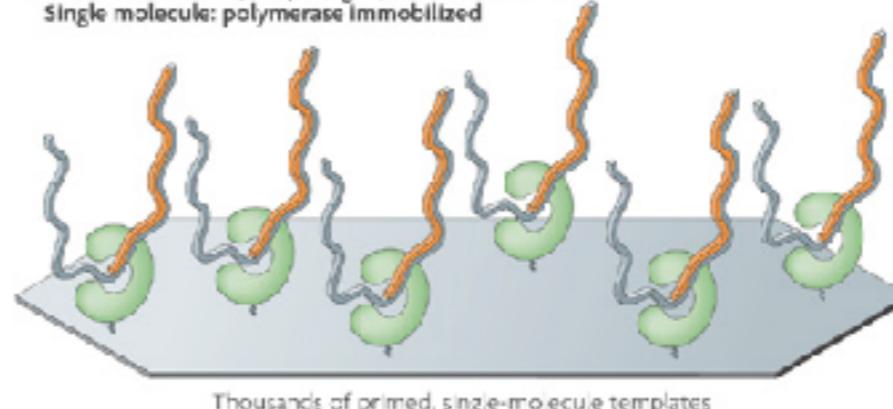
c Helicos BioSciences: one-pass sequencing
Single molecule: primer immobilized



d Helicos BioSciences: two-pass sequencing
Single molecule: template immobilized



e Pacific Biosciences, Life/Visigen, LI-COR Biosciences
Single molecule: polymerase immobilized



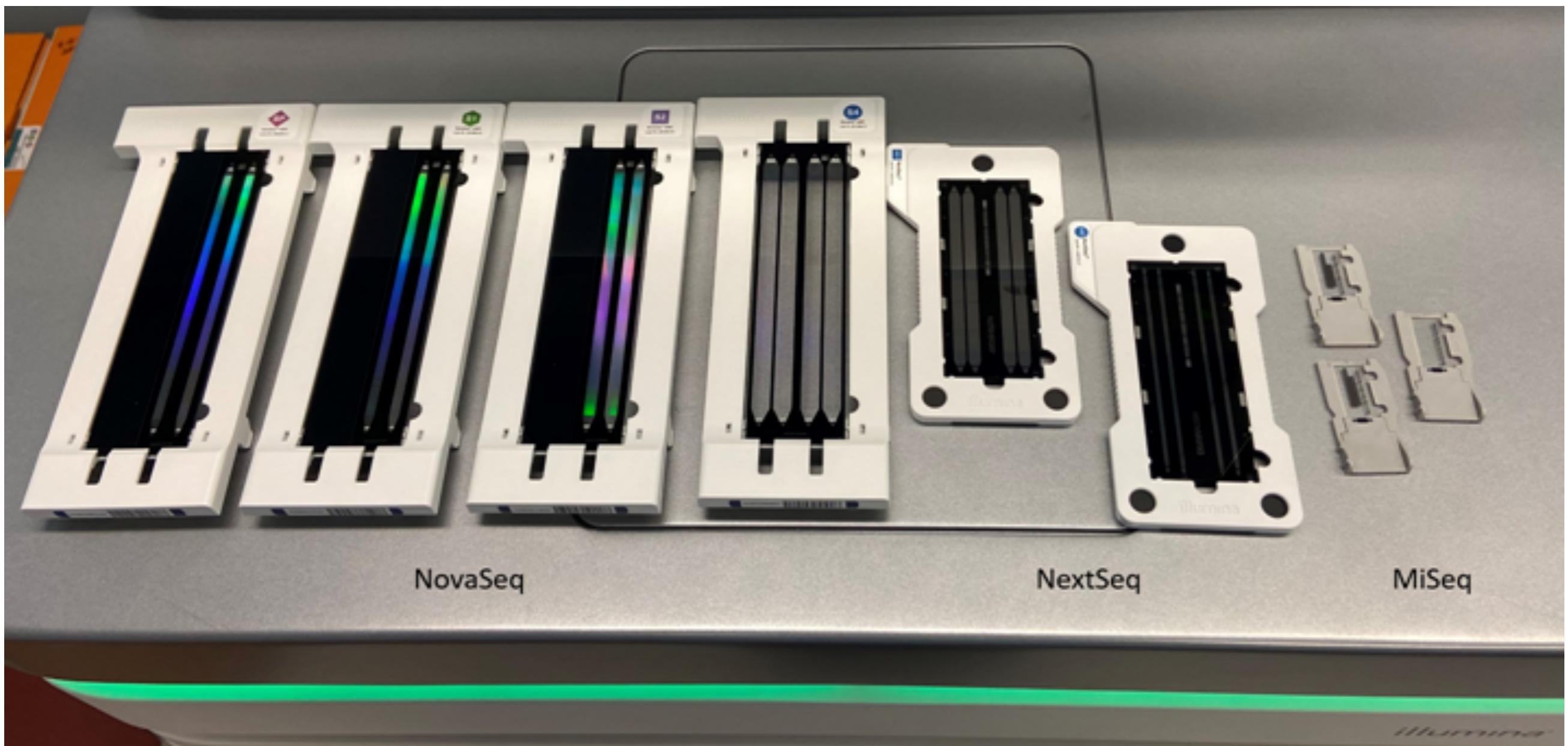
A Flow Cell



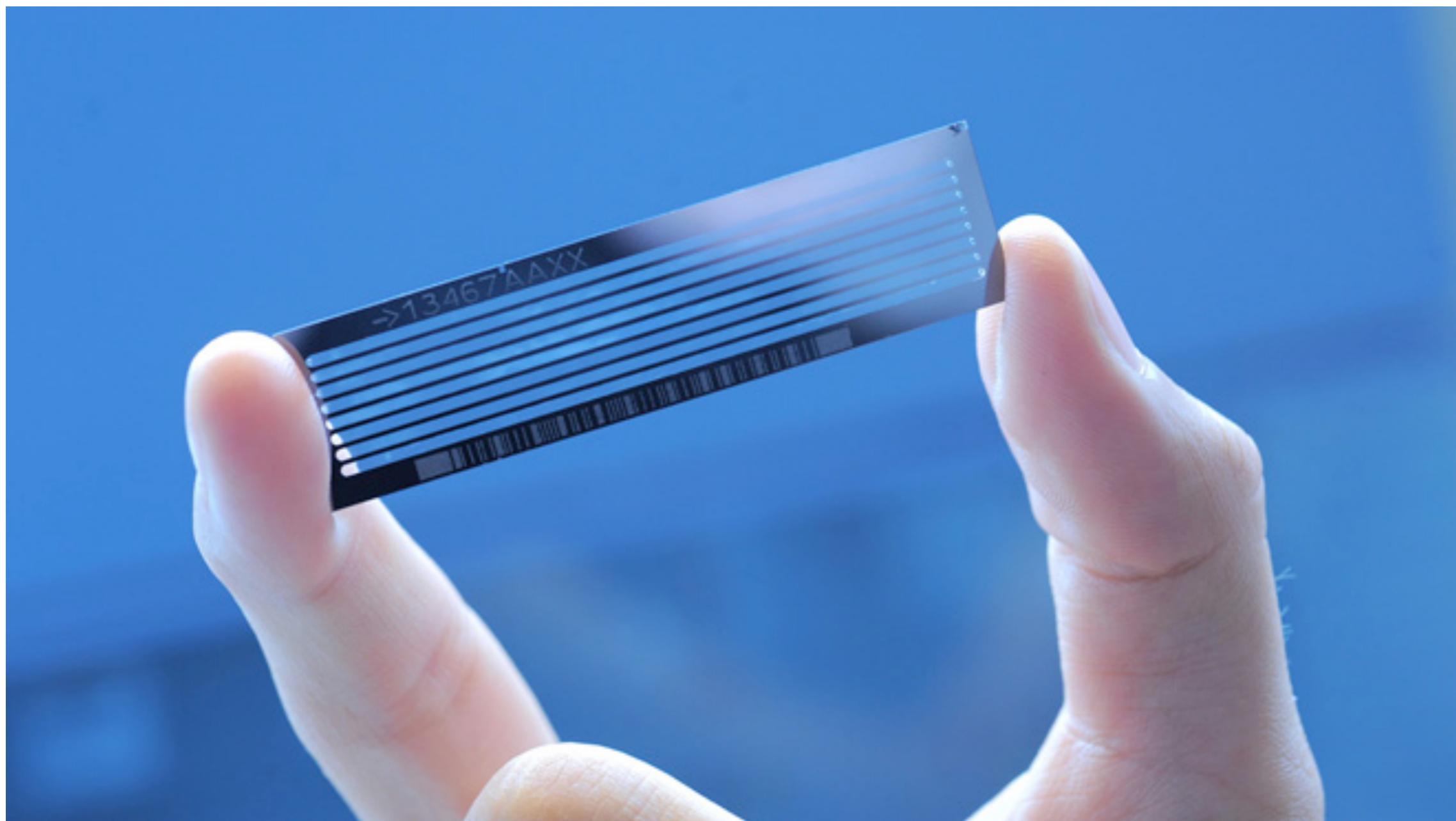
<https://www.lhsc.on.ca/palm/img/hcp.jpg>

<https://www.illumina.com/company/news-center/multimedia-images.html>

A Flow Cell



A Flow Cell



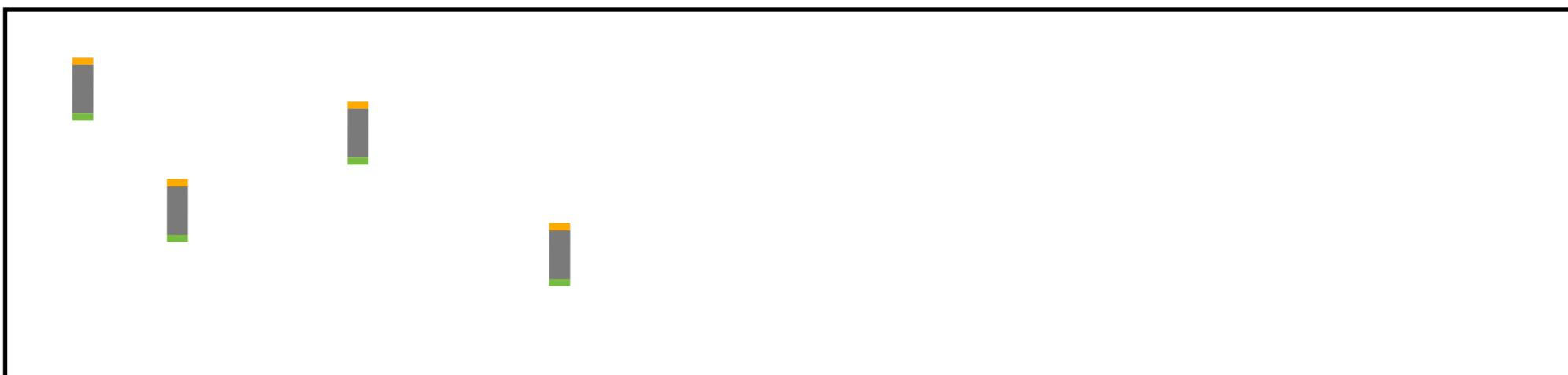
SBS: Sequencing by Synthesis

An Illumina Story

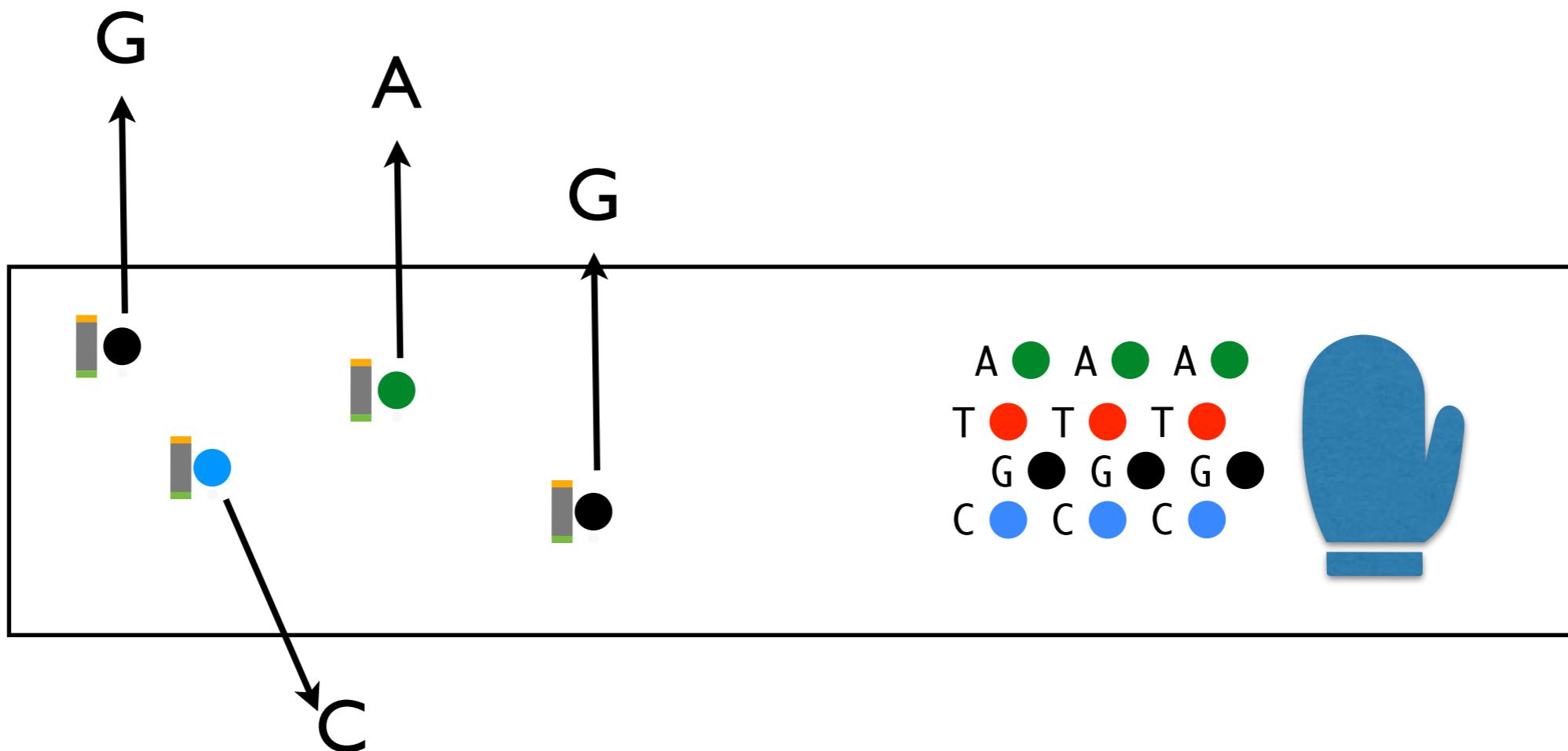
A Flow Cell



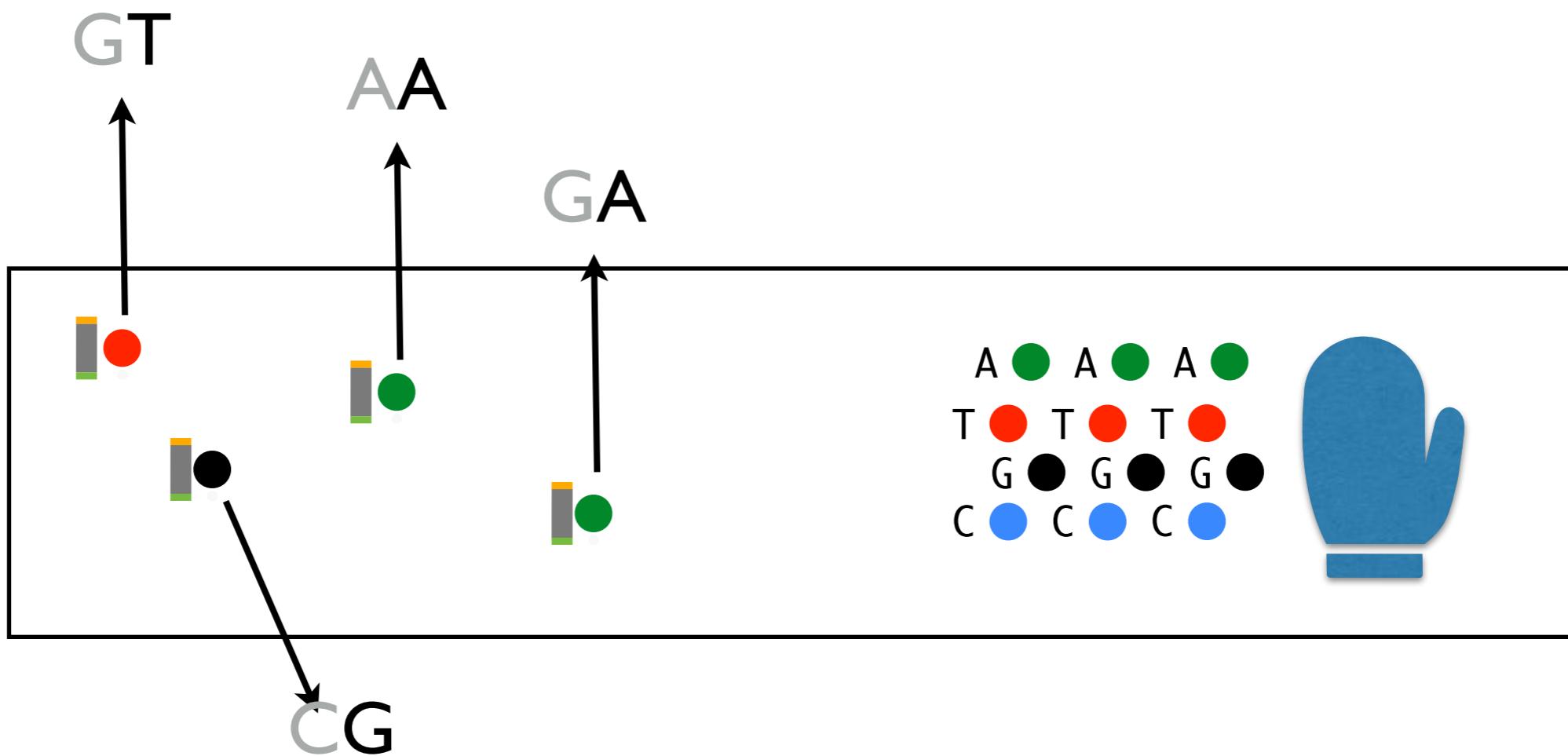
Bind Library



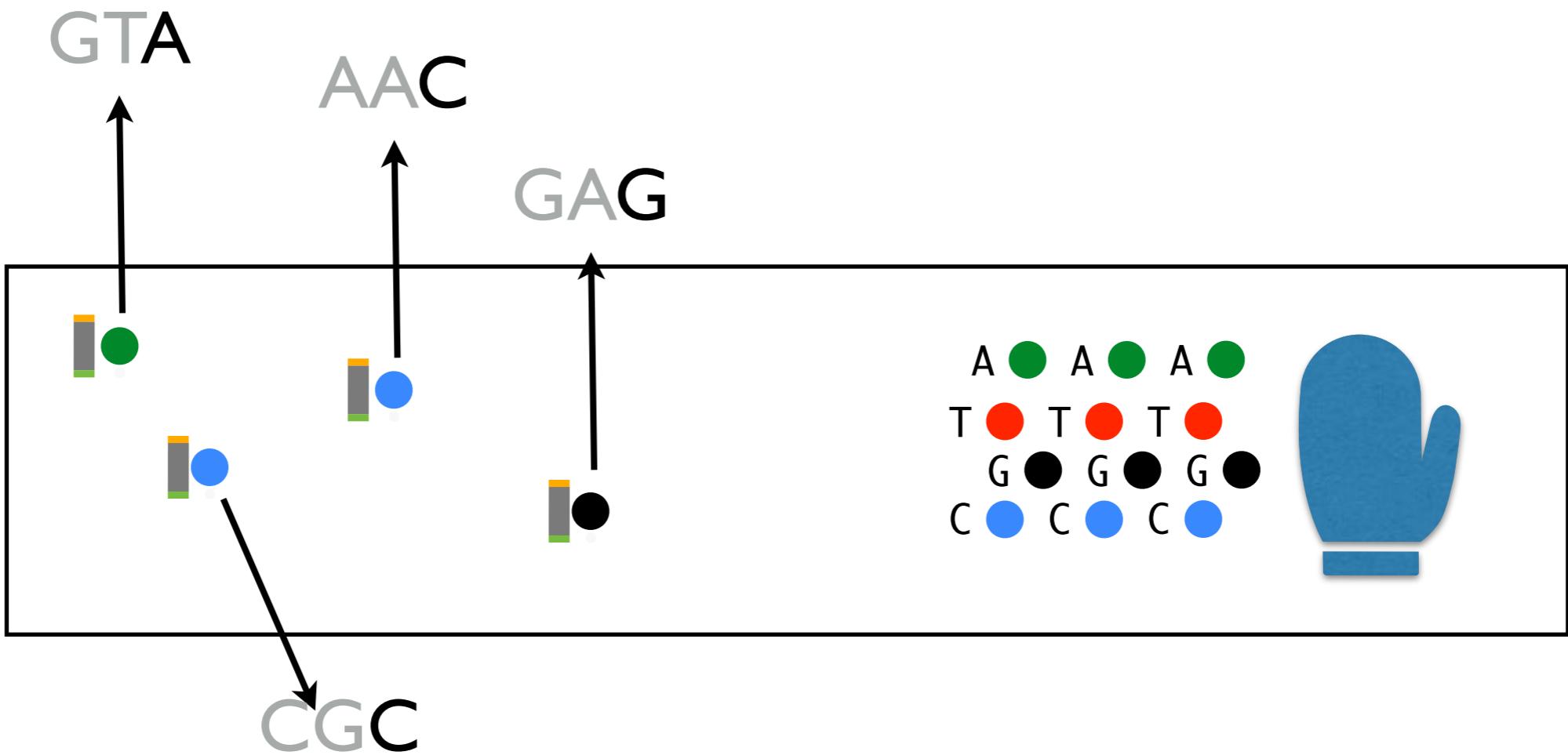
1st Cycle



2nd Cycle



3rd Cycle

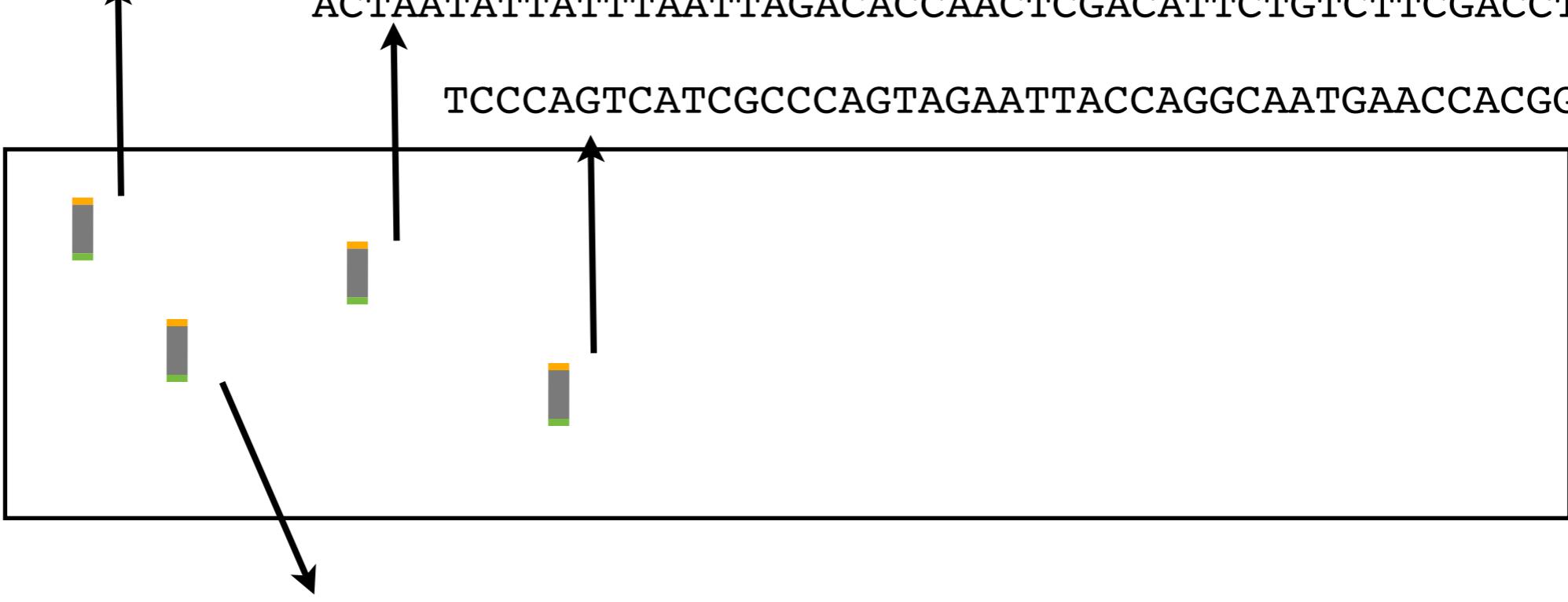


50th Cycle

GAATTCTAAAACAGTTGCATTCTATAATTACAAAATAATTGAAACACTTC

ACTAATATTATTAAATTAGACACCAACTCGACATTCTGTCTTCGACCTAT

TCCCAGTCATCGCCCCAGTAGAATTACCAGGCAATGAACCACGGCCTTCA

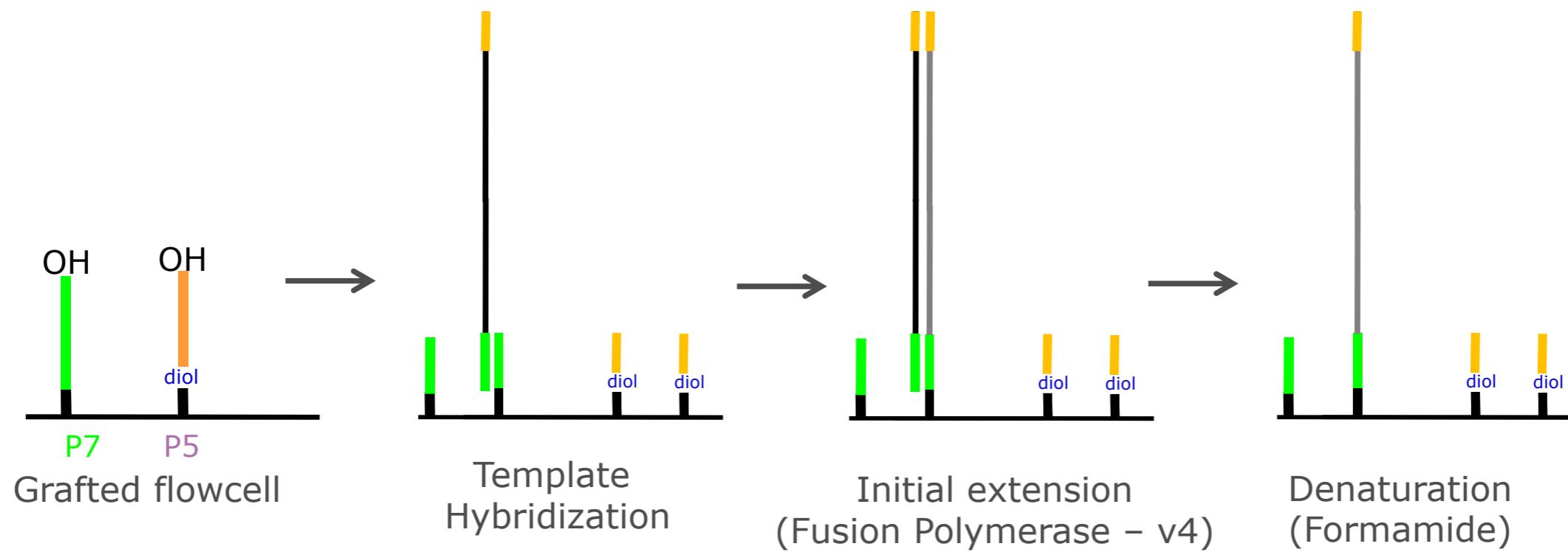


ACAGCATATGGGTTCACTCCAACAGTGAACCATTCCAAAAGGCCTTGCCT

Illumina Short Reads

- 50 - 300bp

Cluster generation – hybridization and amplification



Hybridization

5' -CTGATCTGACTGATGCGTATGCTAGT-3'

+

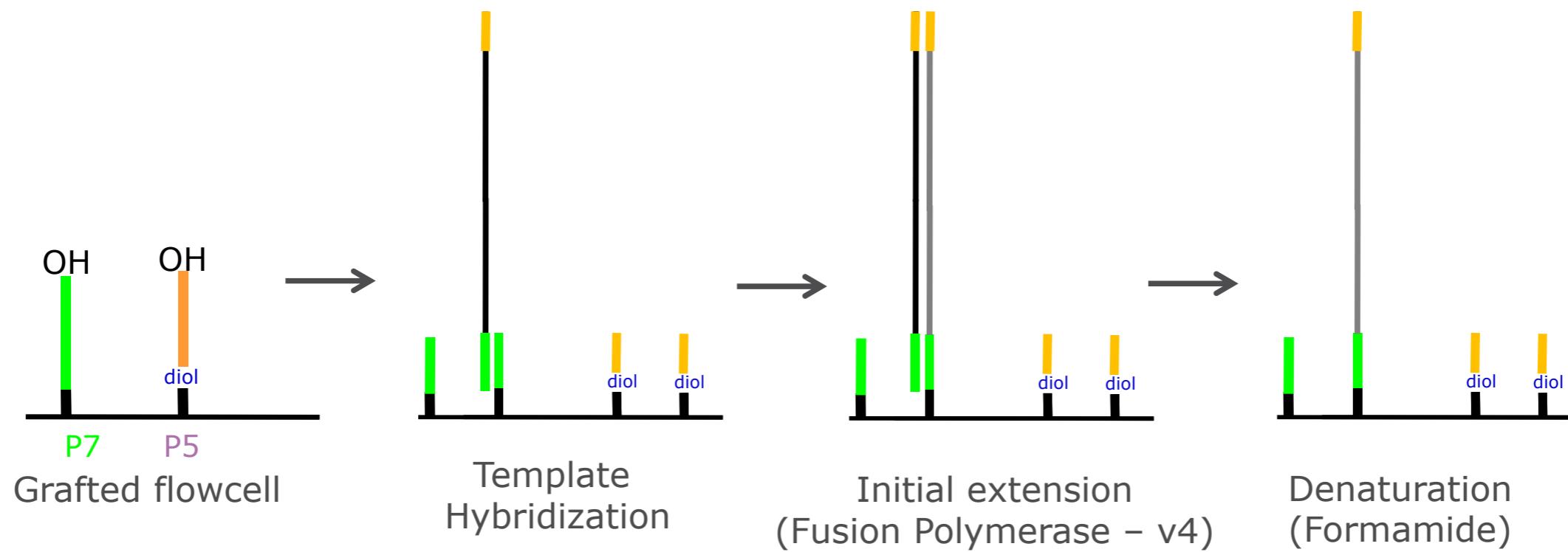
3' -**GCATAC**-5'

=

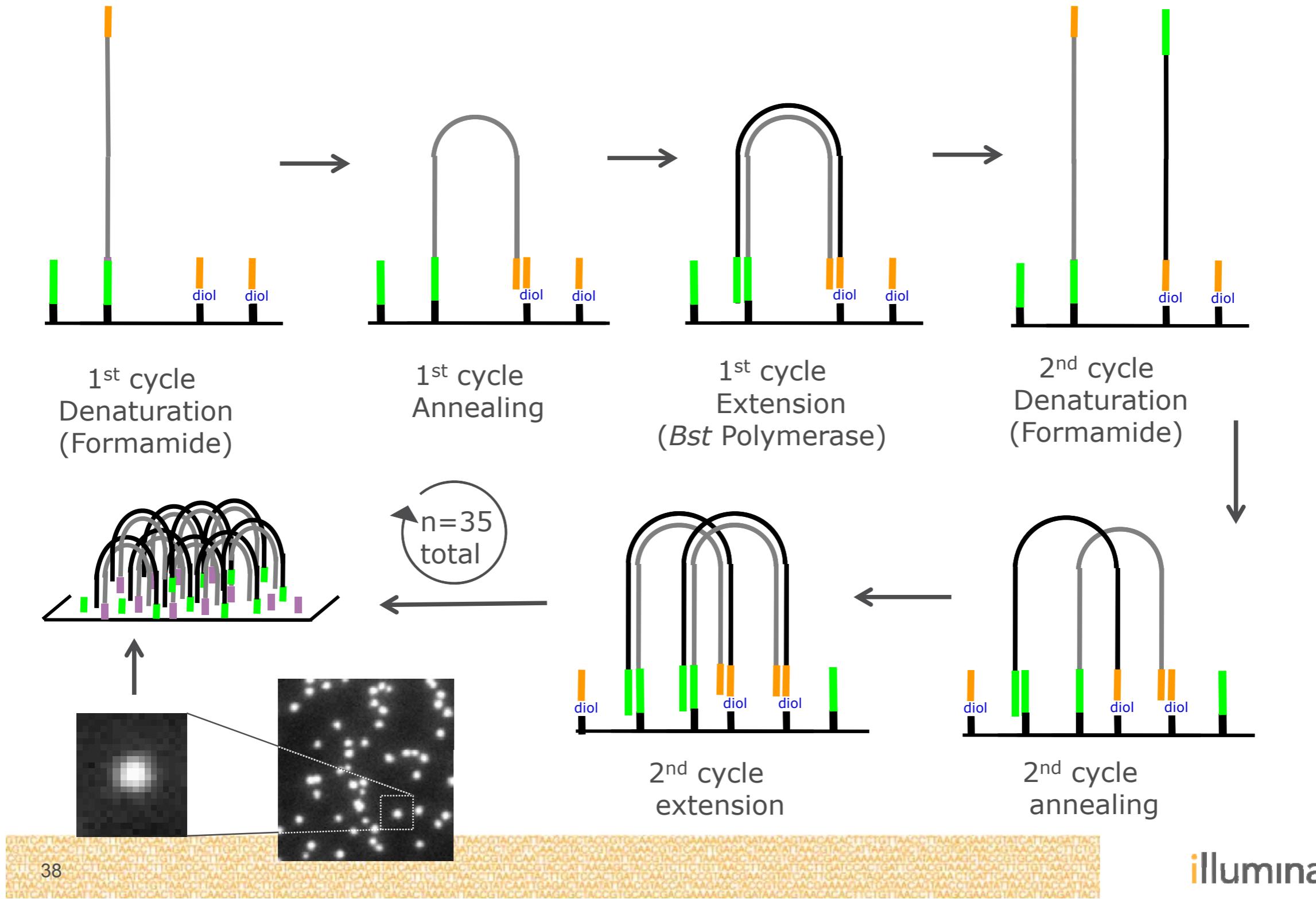
5' -CTGATCTGACTGATGCGTATGCTAGT-3'

3' -**GCATAC**-5'

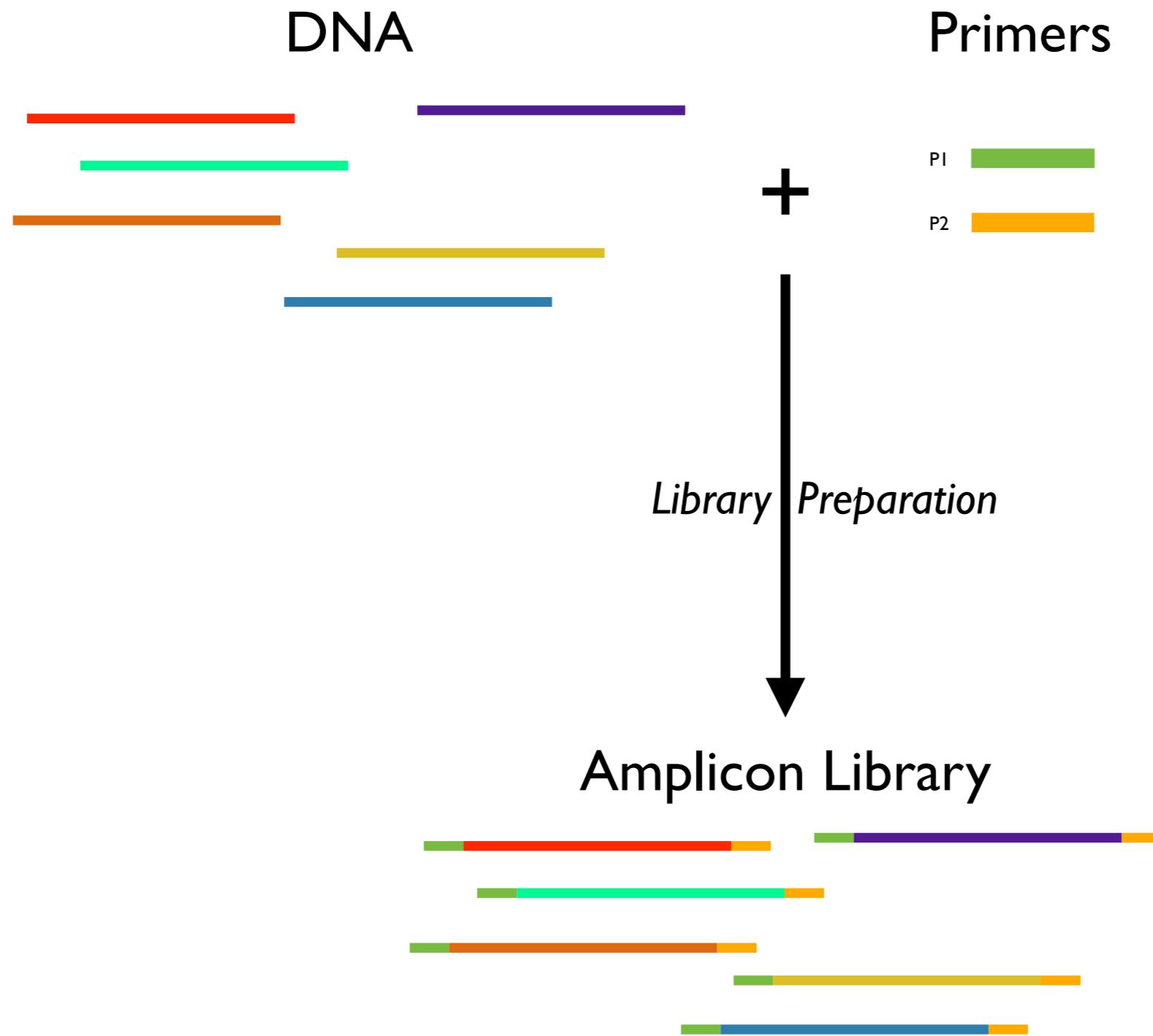
Cluster generation – hybridization and amplification



Cluster generation – hybridization and amplification



Library Preparation

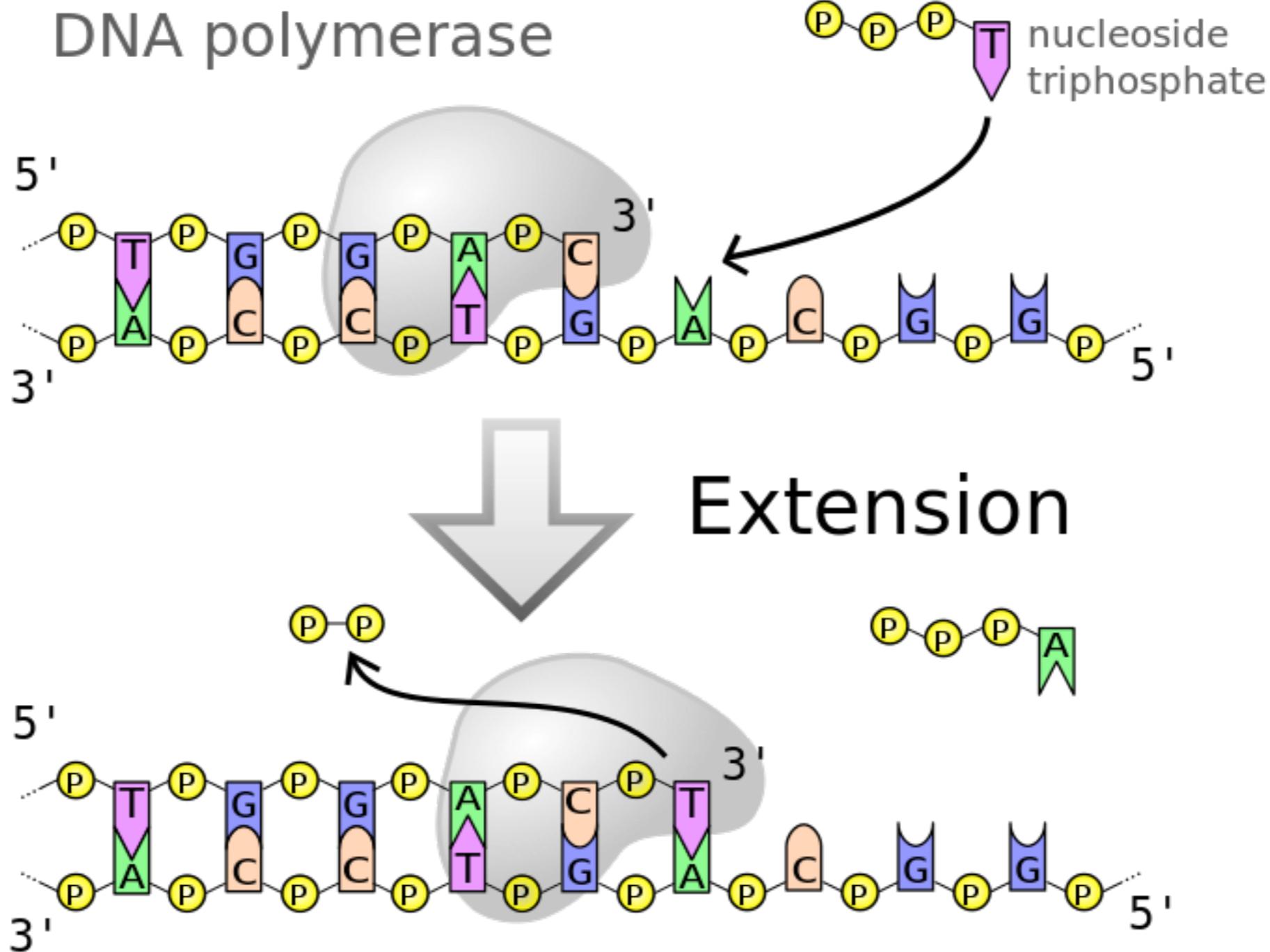


Why Adapters?

DNA Synthesis

- What are the minimum components for DNA Replication?

DNA Synthesis

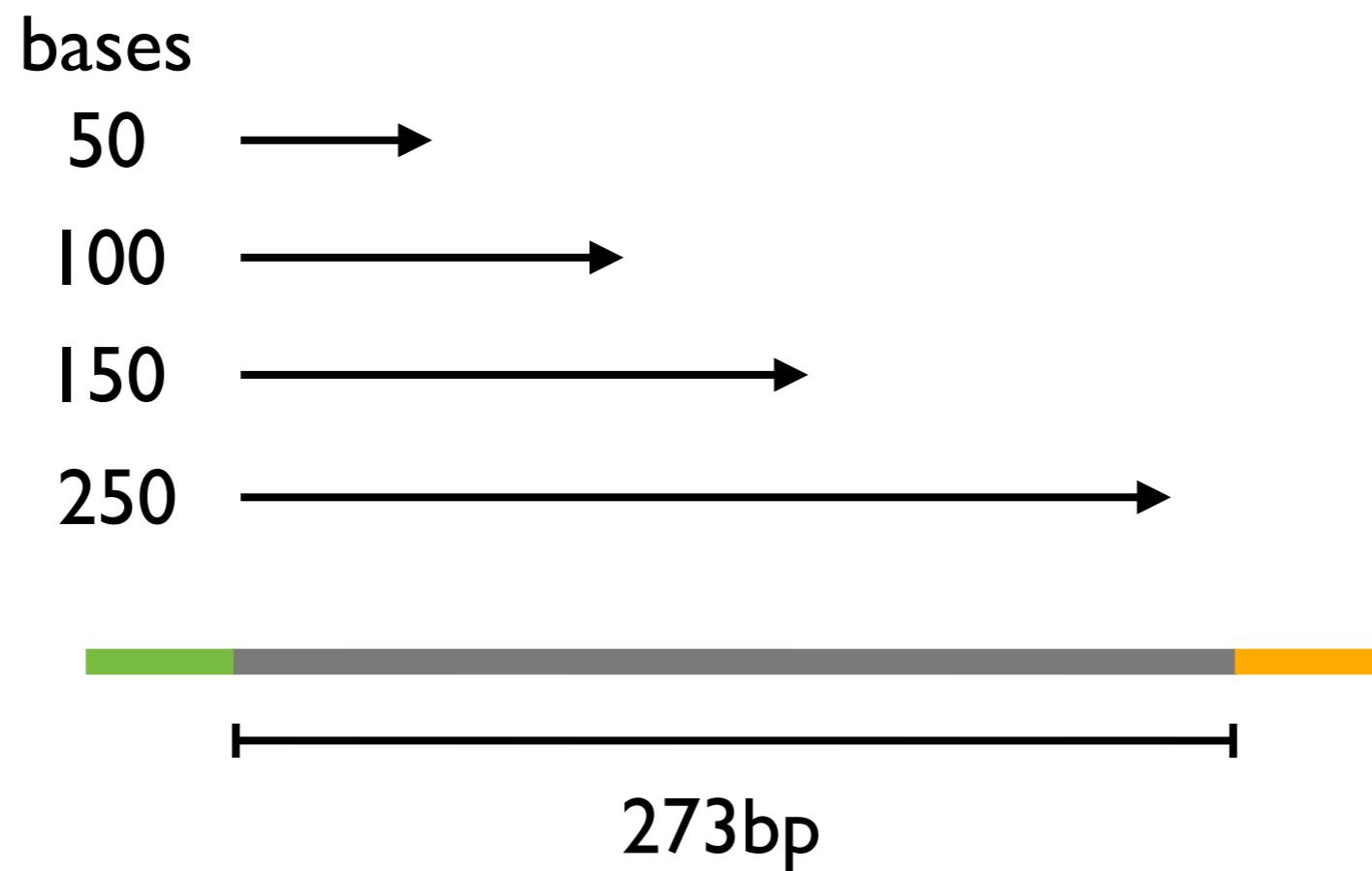


Why Adapters?

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

Additional Sequencing Details

Read Length

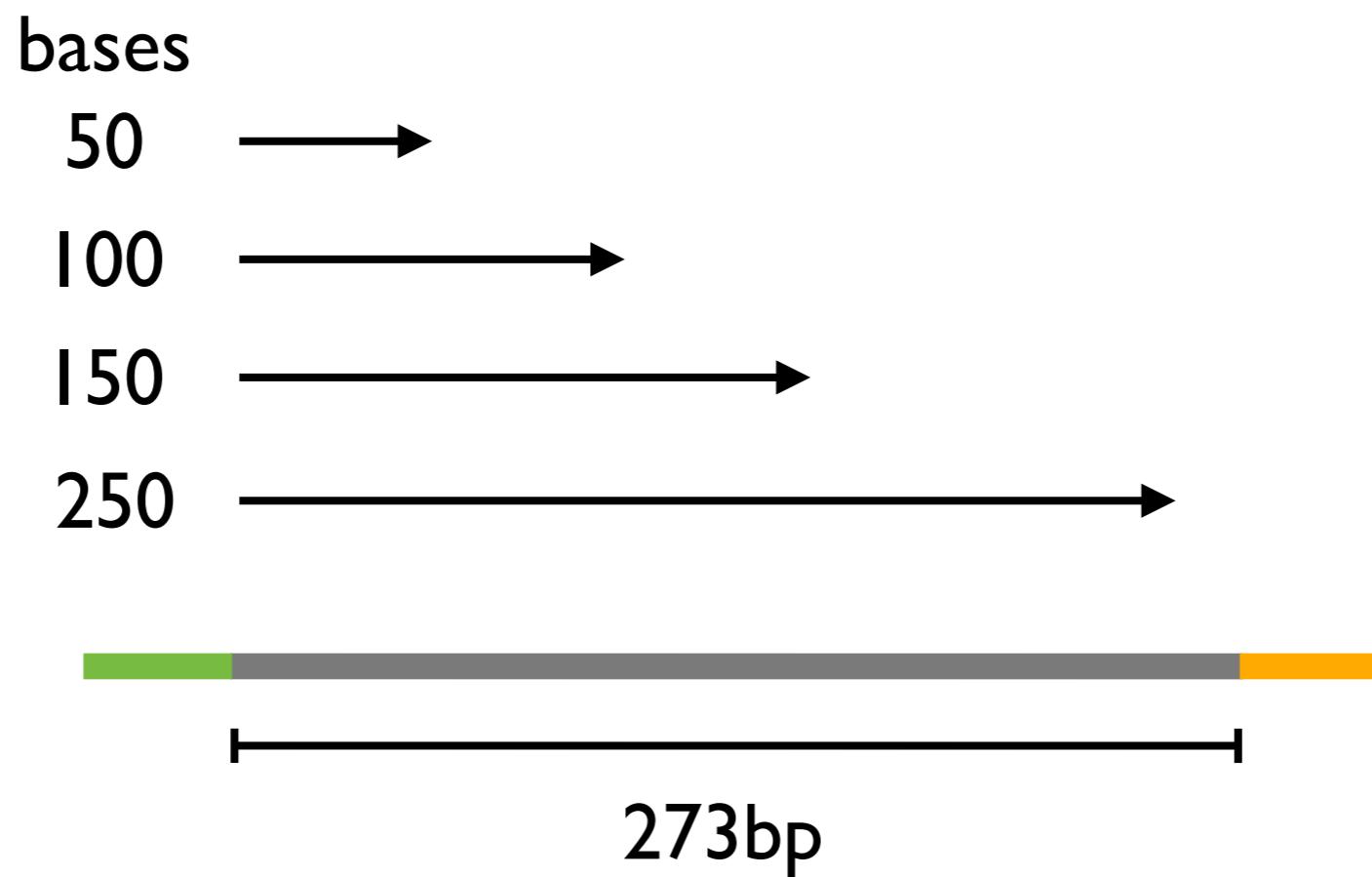


Paired-End

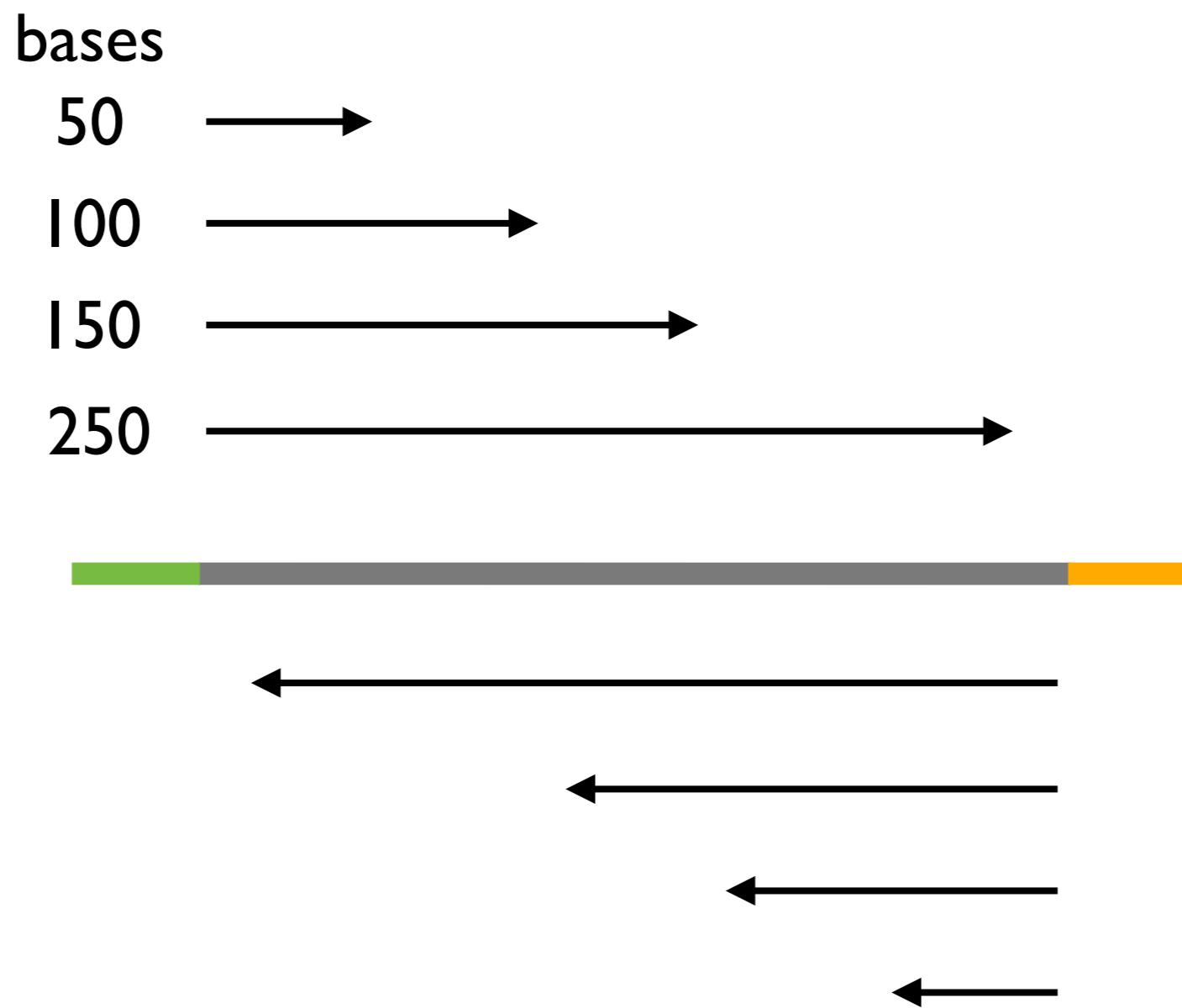
TCGAAAAG
AGCTTTCAATTCTGACTGCAACGGGCAATATGTCTGTGTGGA

AGCTTTCAATTCTGACTGCAACGGGCAATATGTCTGTGTGGA
GACACACCT

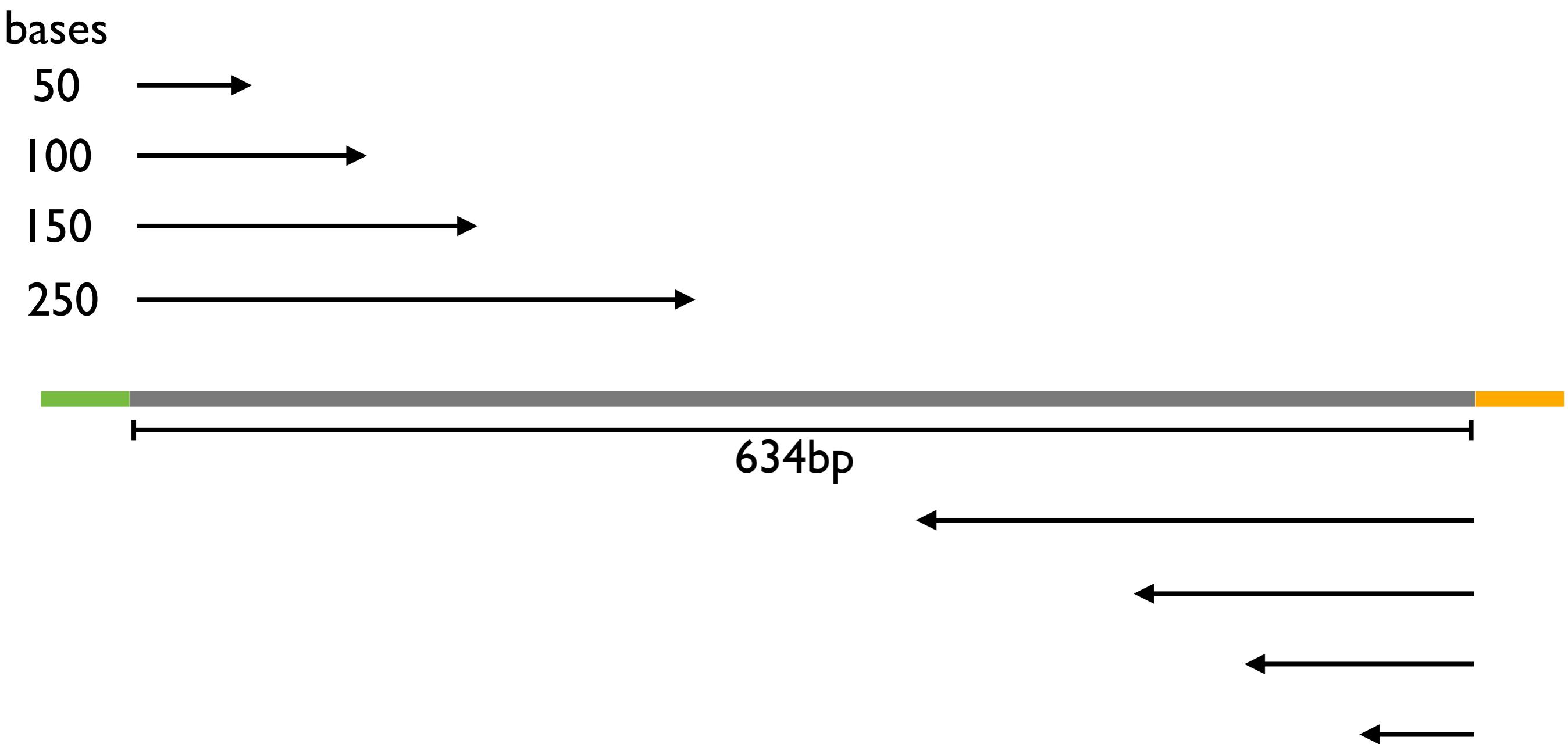
Read Length



Read Length

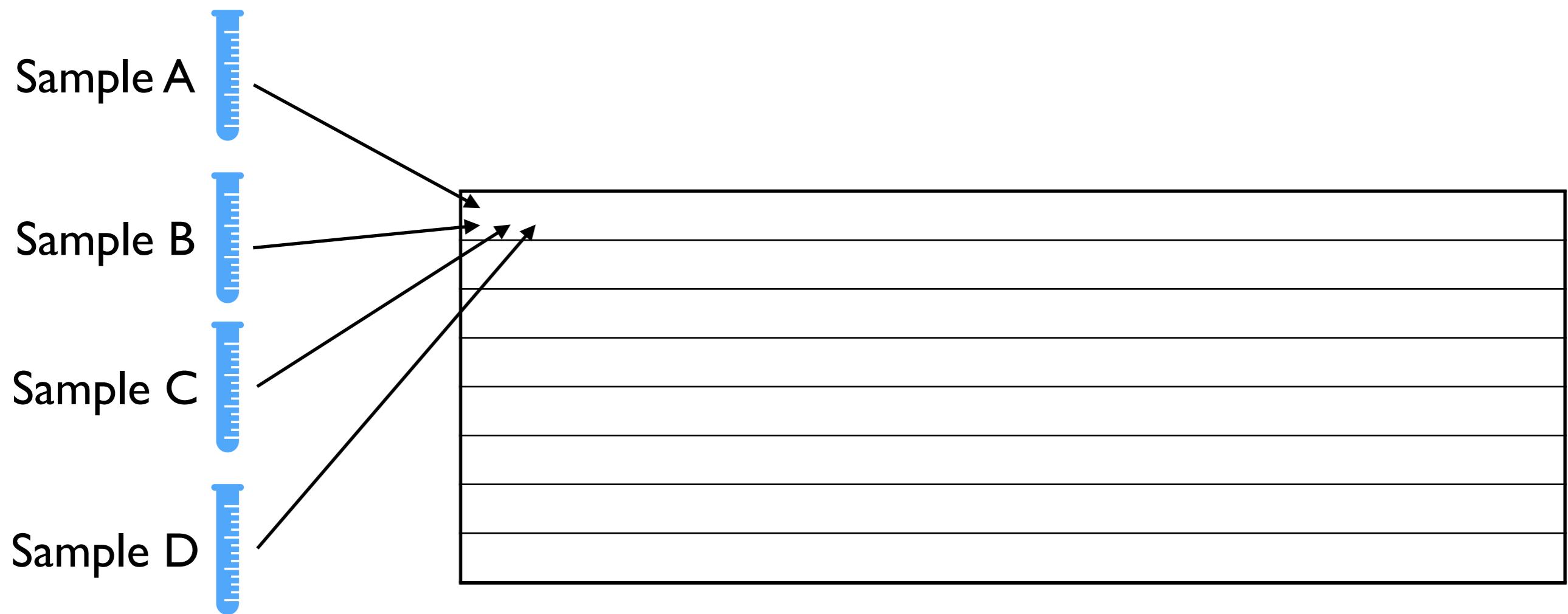


Read Length

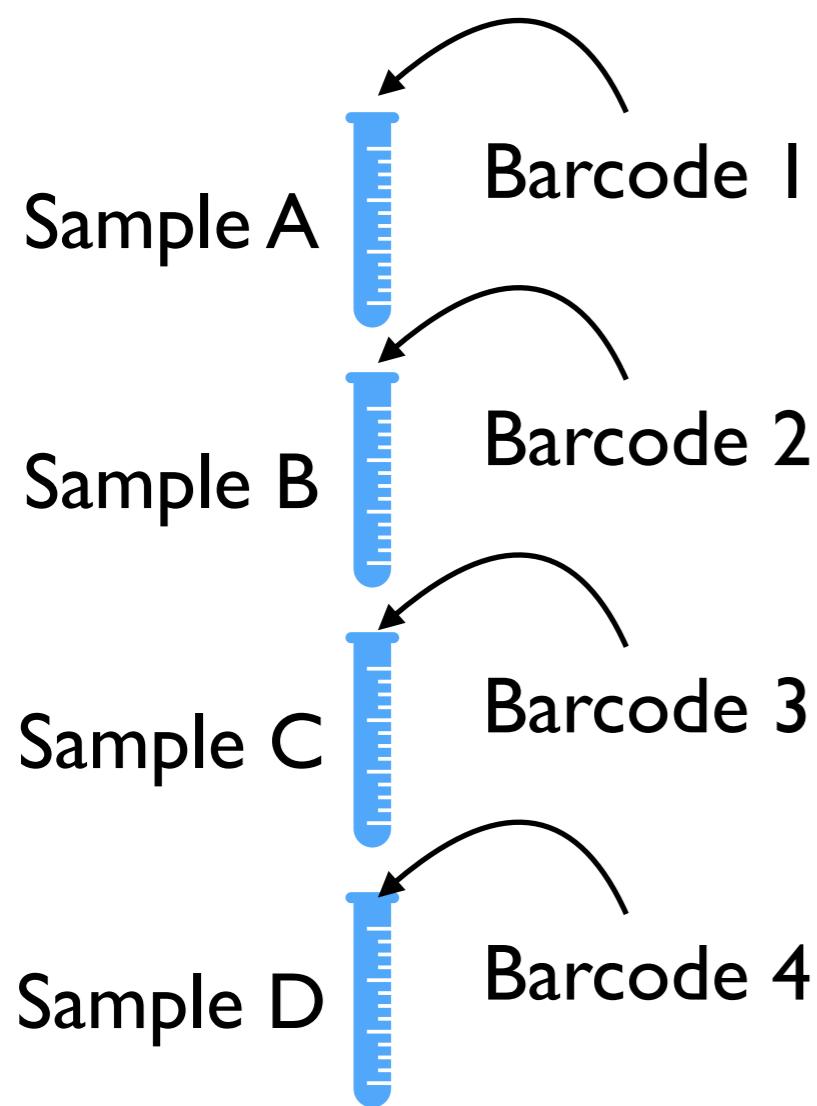


Barcodes: Why?

Multiplexing: Combine multiple samples in a lane



Barcodes

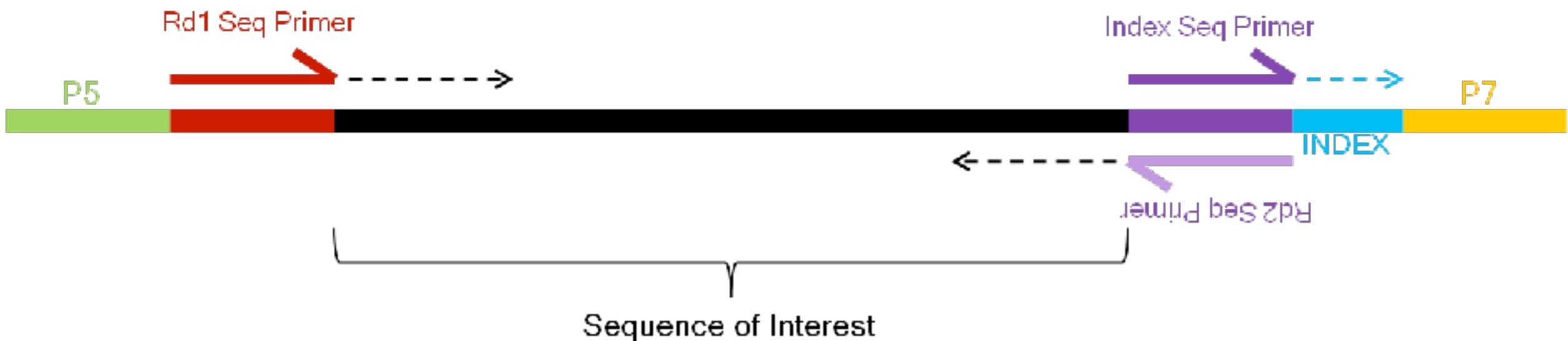


Barcodes

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	CTCGAACCA
10_B	P58-E10	CAACTCCA

Multiplexing (Barcodes)

STRUCTURE DETAILS



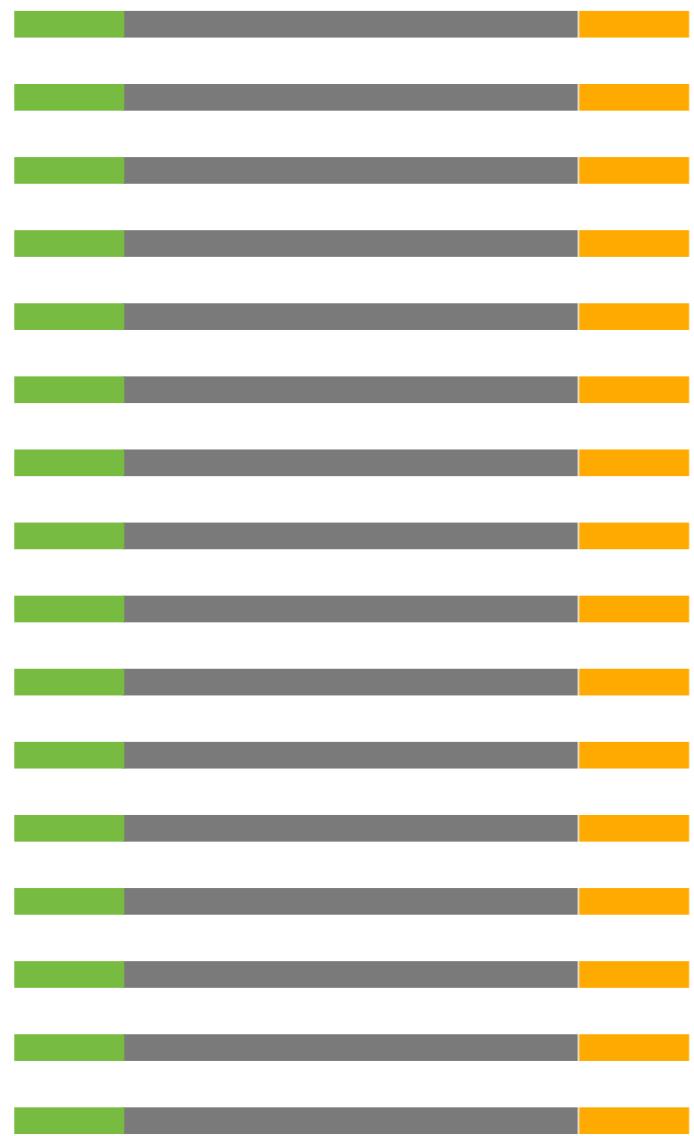
MiSeq, NextSeq, and More Seqs

	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 × 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

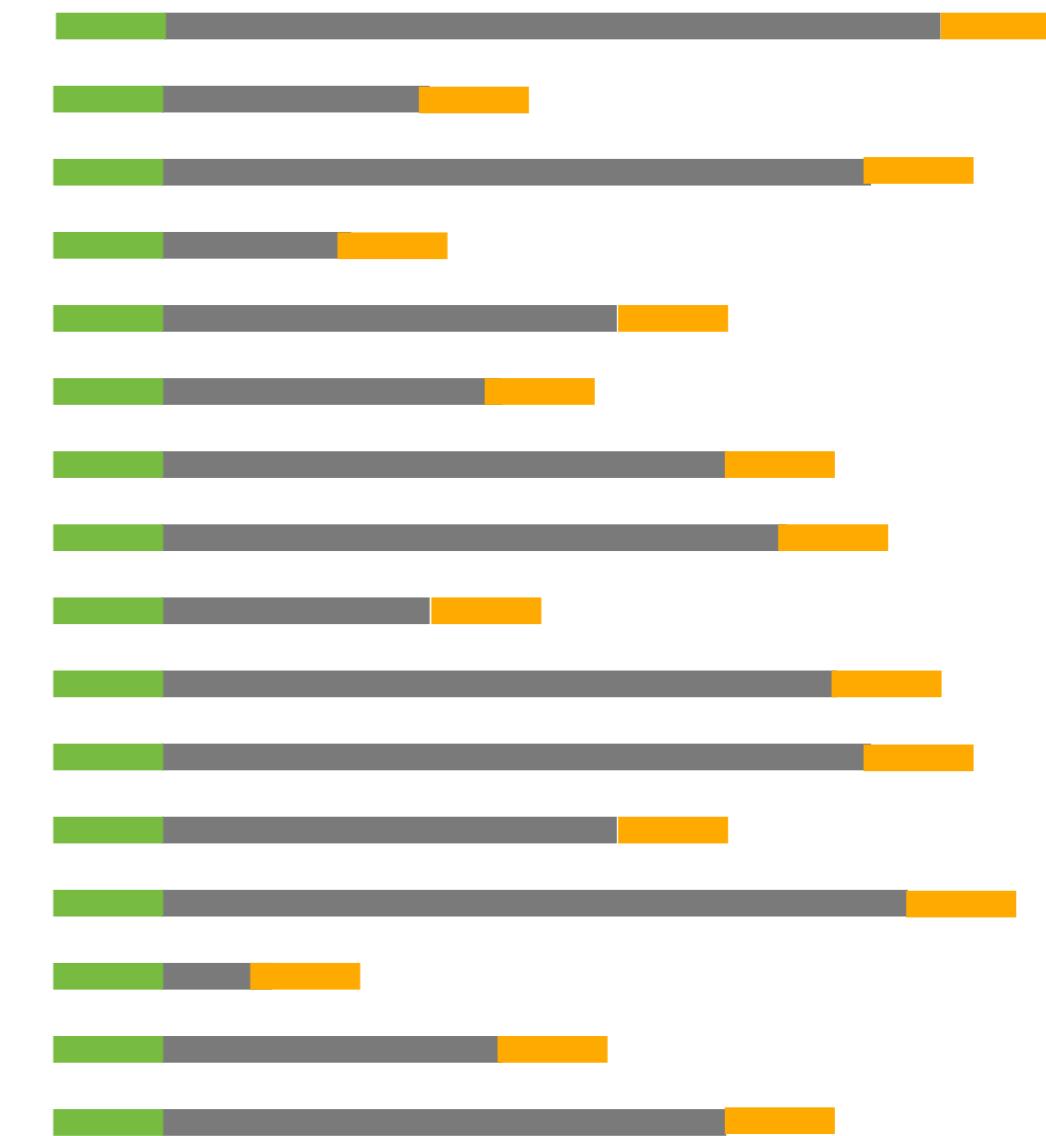
* Duke Sequencing and Genomic Technologies Shared Resource, July 2018

Sequencing Library

Amplicon Library



Shotgun Library



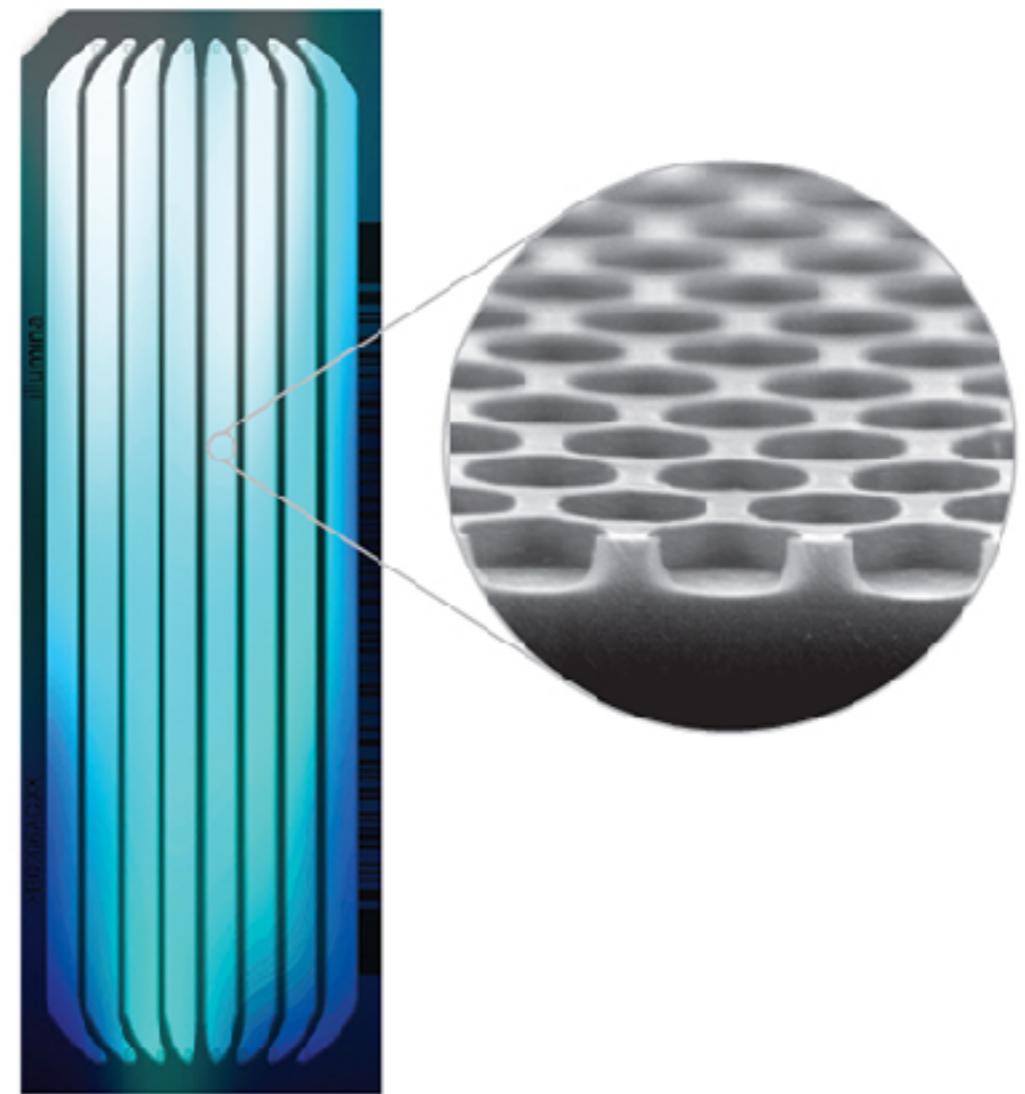
1st Generation	2nd Generation	3rd Generation
Chemical (Maxim-Gilbert)	Pyrosequencing (454)	Single molecule real time (PacBio)
Chain Termination (Sanger)	Chain Termination (Illumina)	Nanopore sequencing (Oxford Nanopore)
Pyrosequencing	Sequencing by ligation (SOLiD sequencing)	
	Ion semiconductor (Ion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

Illumina Video

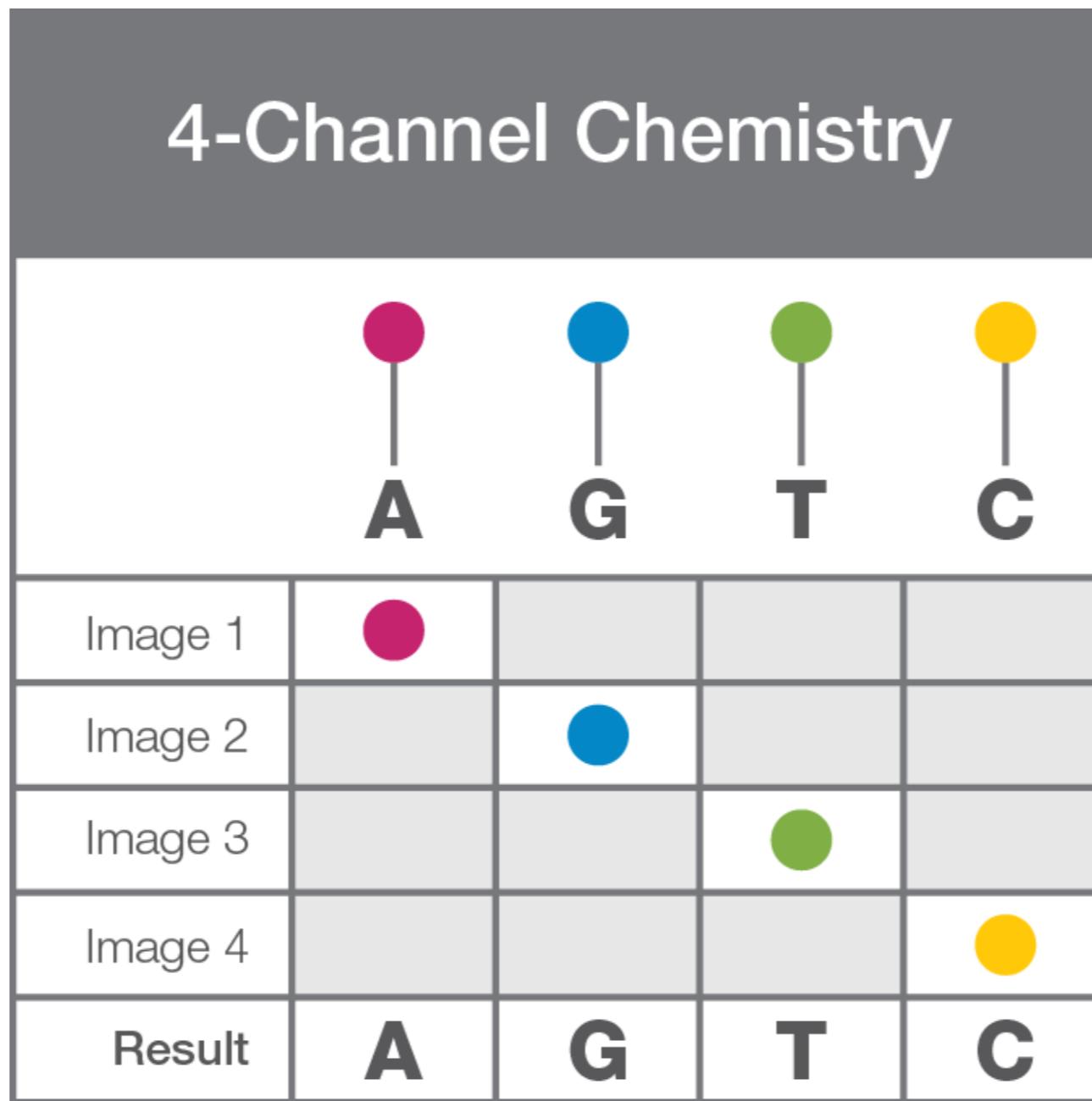
<https://www.youtube.com/watch?v=HMyCqWhwB8E>

Patterned Flow Cells

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



4-Channel Chemistry



2-Channel Chemistry

4-Channel Chemistry

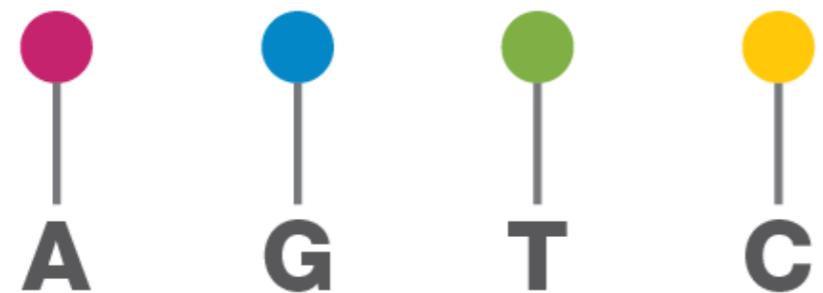


Image 1



Image 2



Image 3



Image 4



Result

A

G

T

C

2-Channel Chemistry



Image 1



Image 2



Result

A

G

T

C