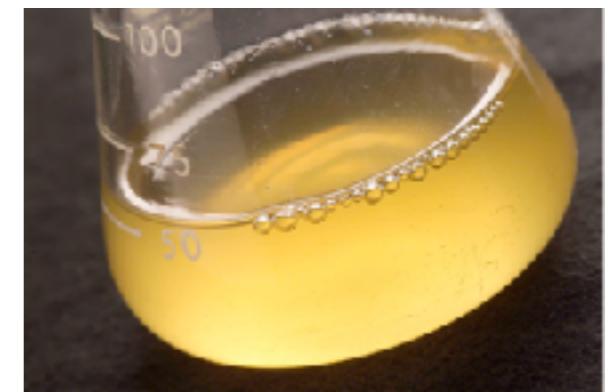


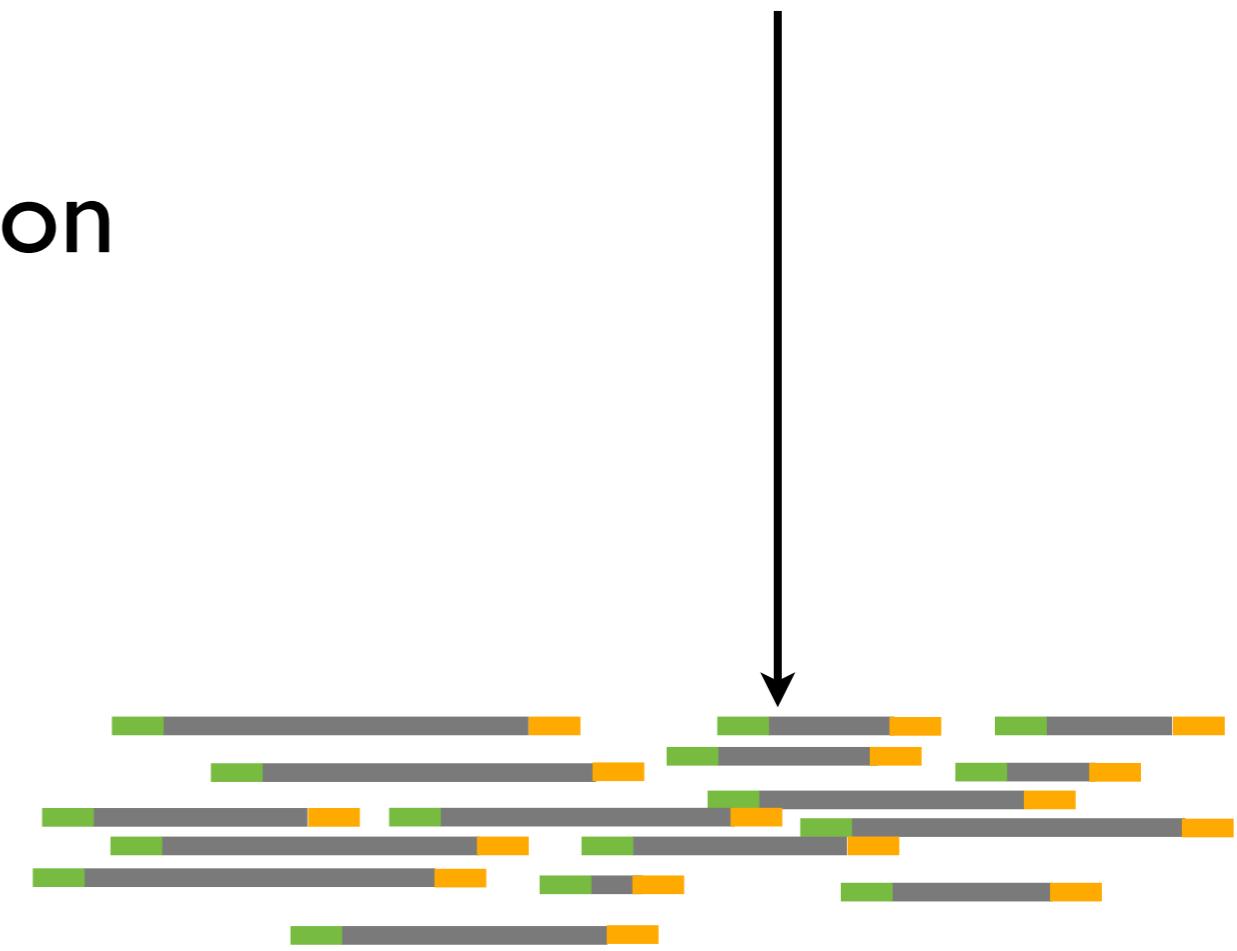
# HTS Background and Theory

Josh Granek

# HTS Experiment: Major Components



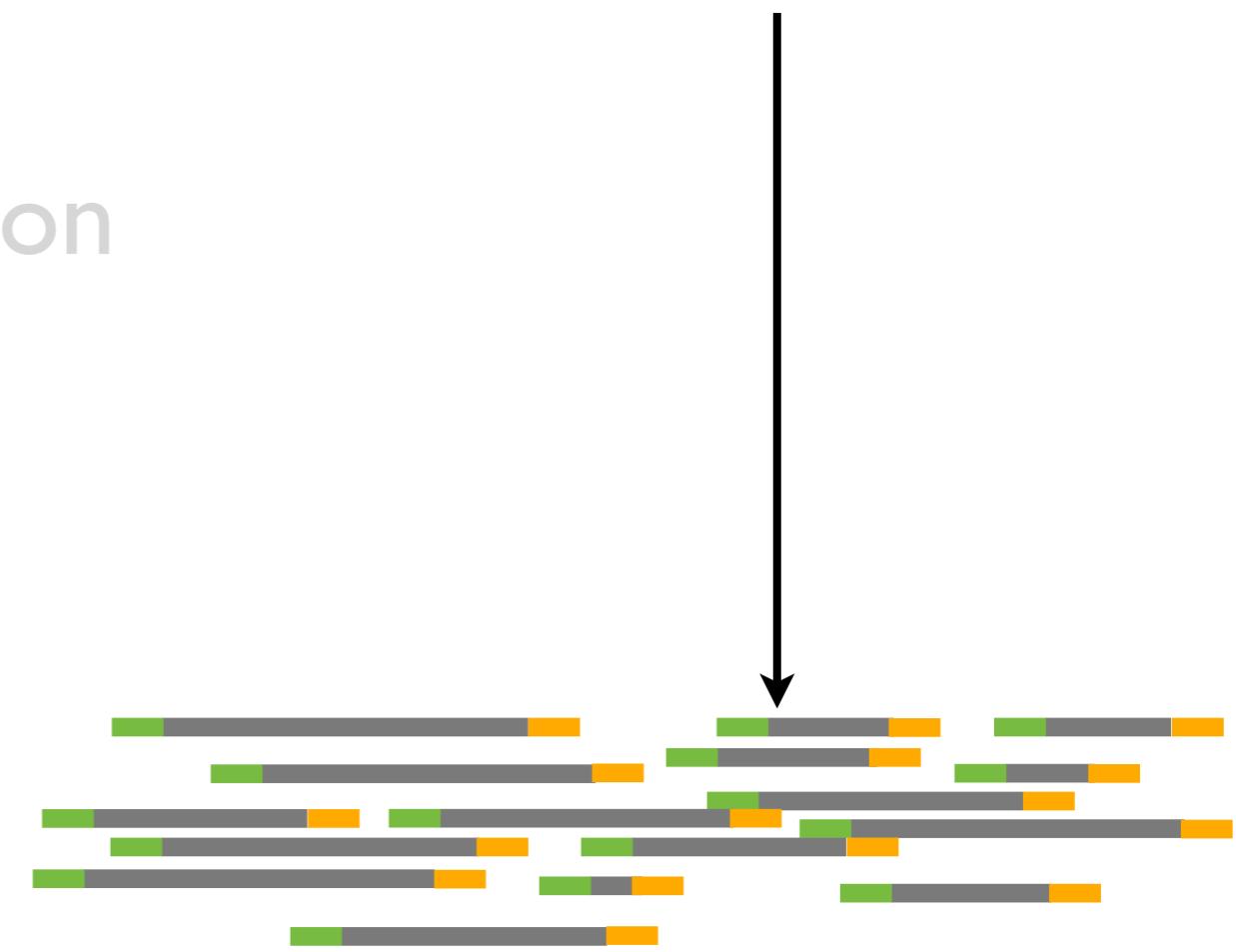
1. Sample Collection
2. Nucleic Acid Extraction
3. Library Preparation



# HTS Experiment: Major Components



1. Sample Collection
2. Nucleic Acid Extraction
3. Library Preparation



# Library Preparation

## Purified Nucleic Acid



Adapter Ligation



# Sanger Sequencing

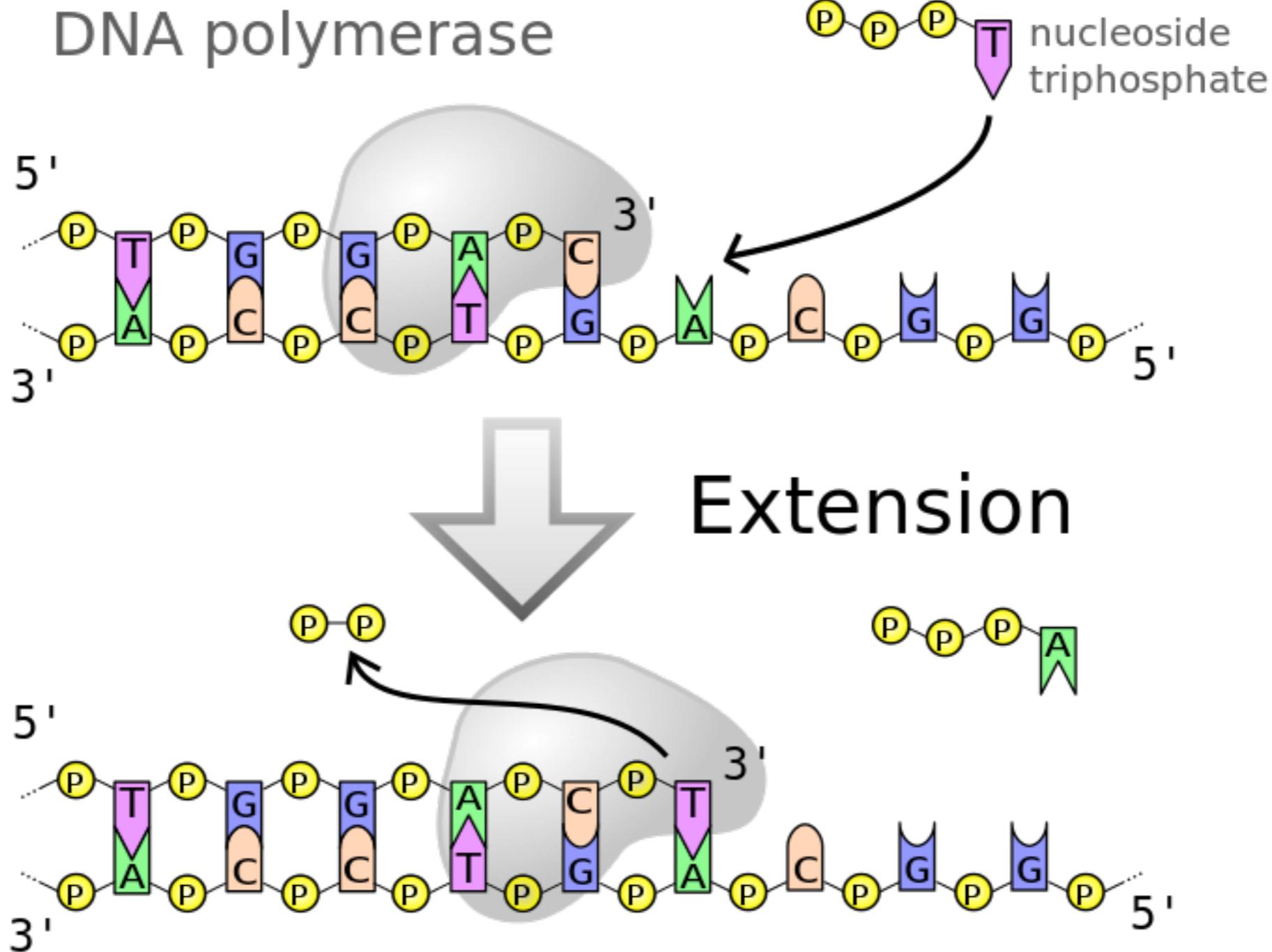
# DNA Synthesis

- What are the minimum components for DNA Replication?

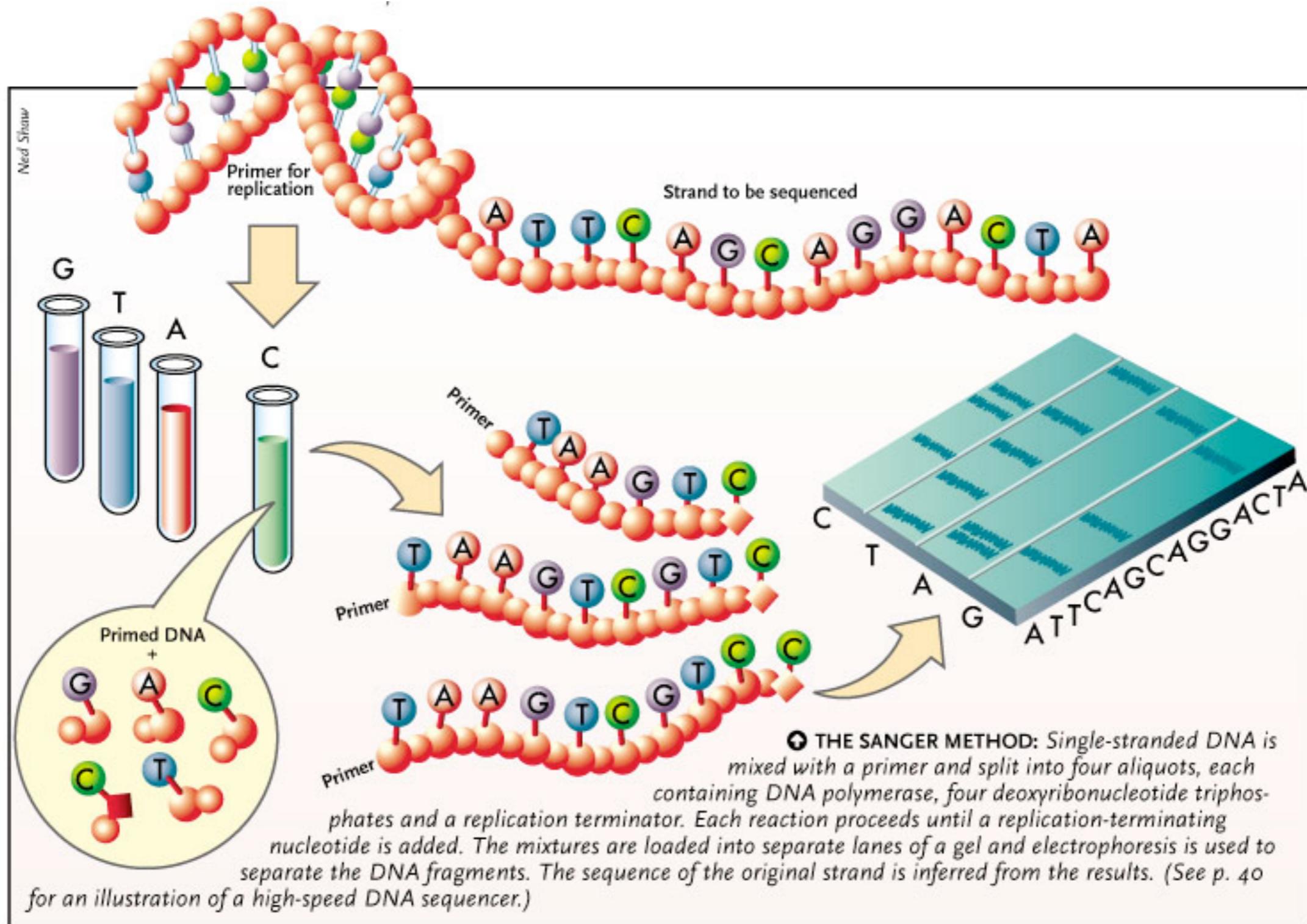
# DNA Synthesis

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

# DNA Synthesis



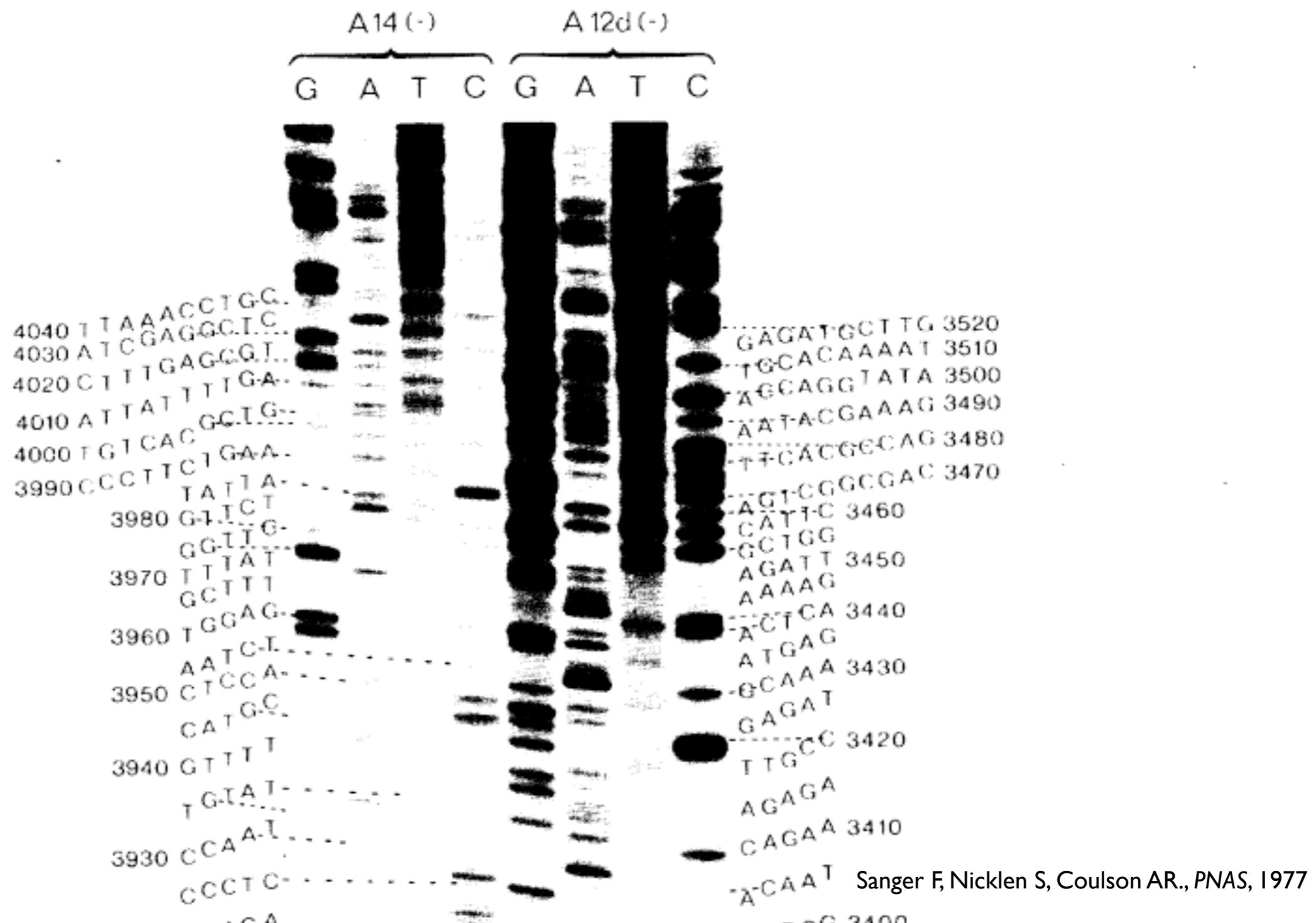
# Sanger Sequencing



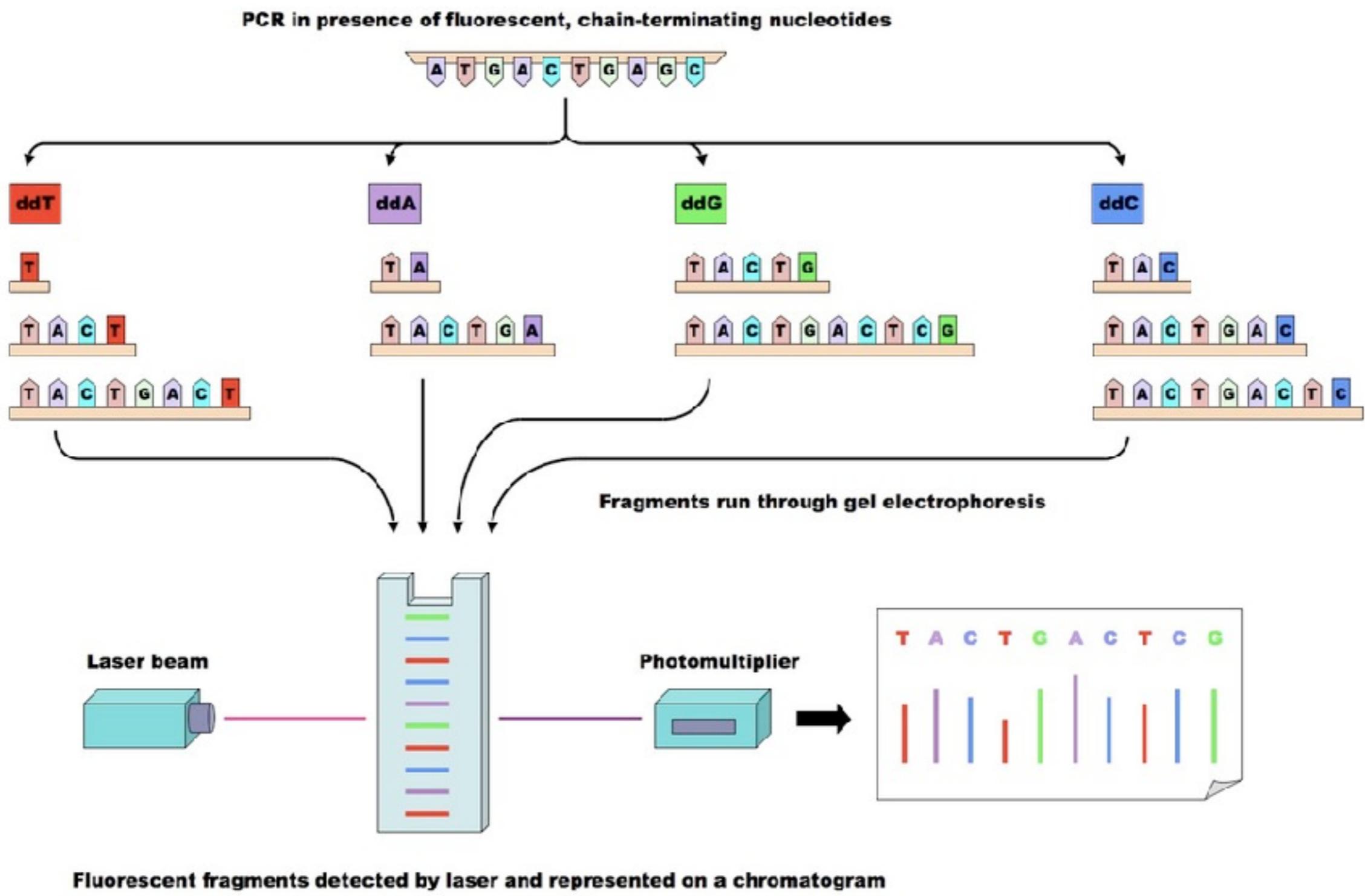
# Sanger Sequencing

5464 Biochemistry: Sanger *et al.*

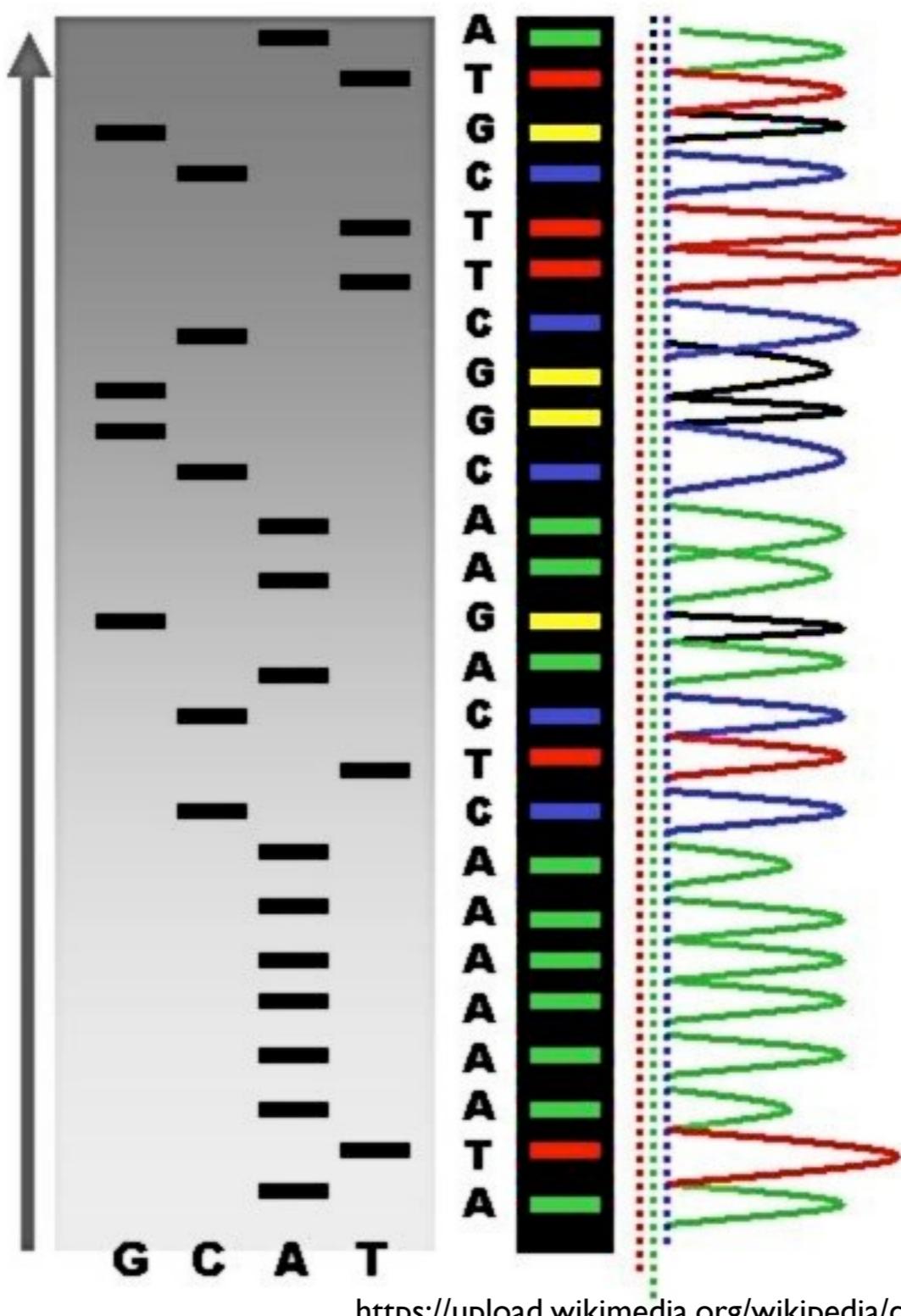
Proc. Natl. Acad. Sci. USA 74 (1977)



# Dye-terminator



# Radiolabel vs. Dye



# High-Throughput Sequencing

# Sequencing



AGCTTTCAATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGA

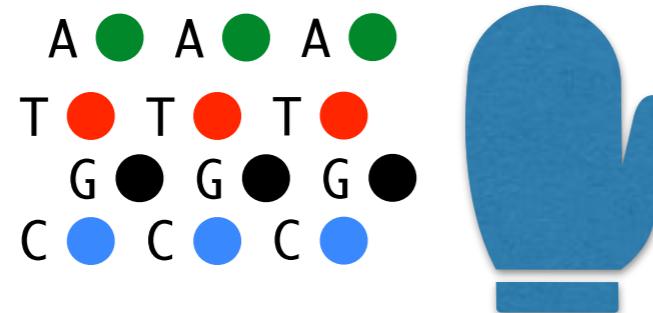
# Sequencing



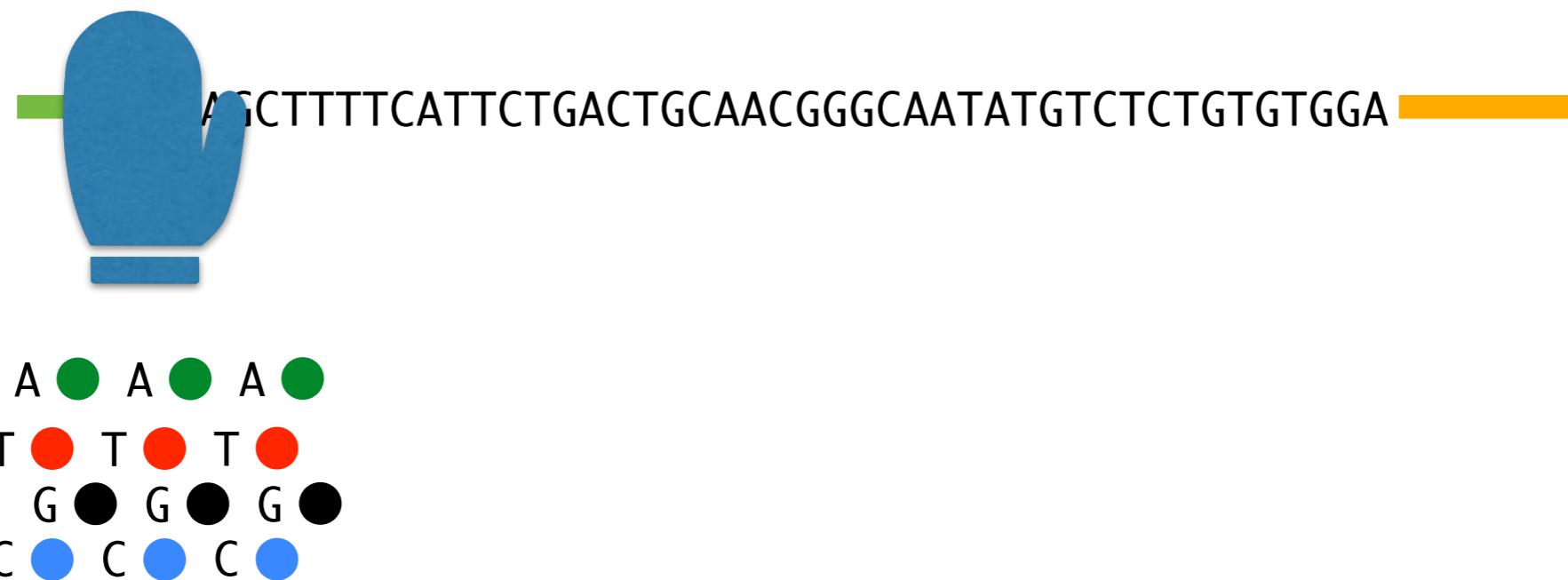
AGCTTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGA

# Sequencing

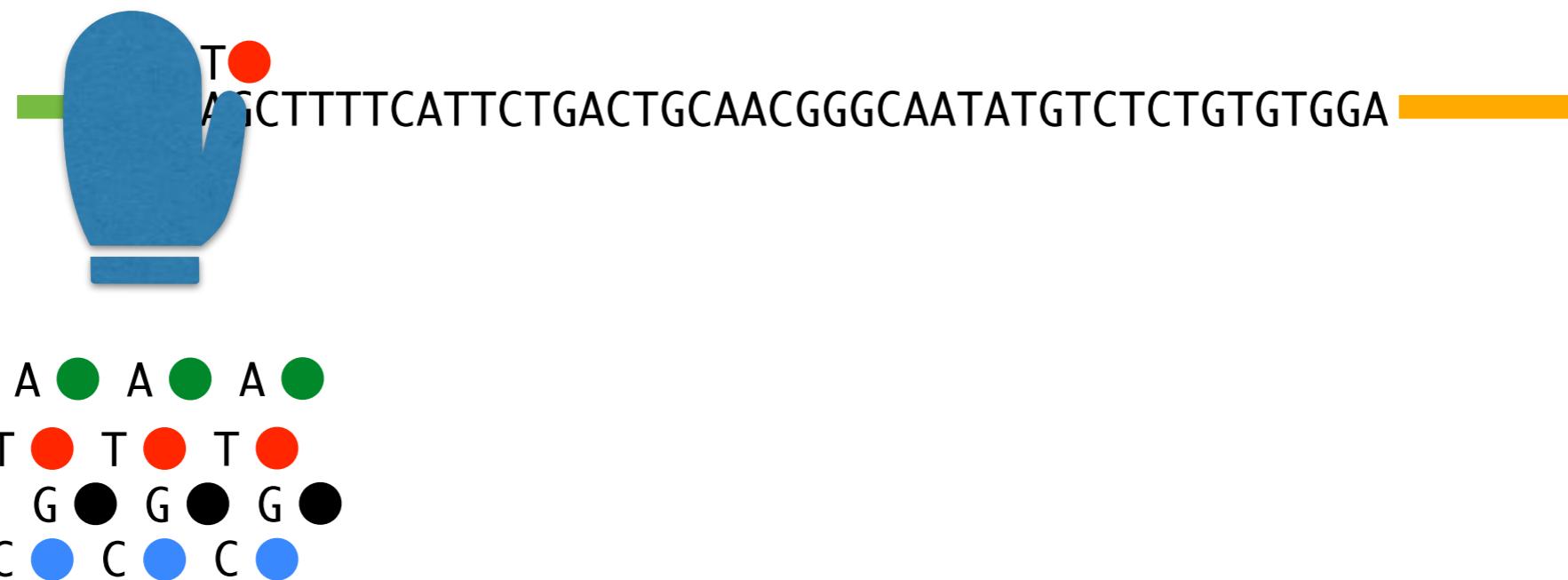
AGCTTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGA



# Sequencing



# Sequencing



# Sequencing

T  
AGCTTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGA

# Sequencing

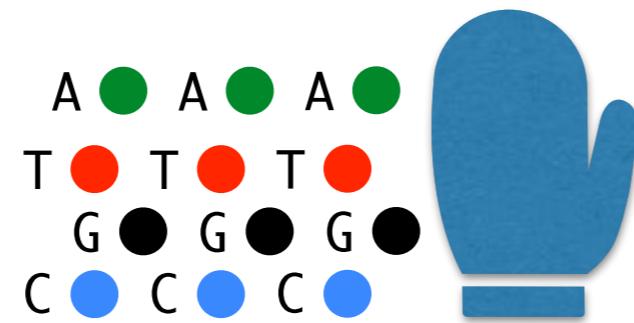
T  
AGCTTTTCATTCTGACTGCAACGGGCAATATGTCTGTGTGGA

# Sequencing

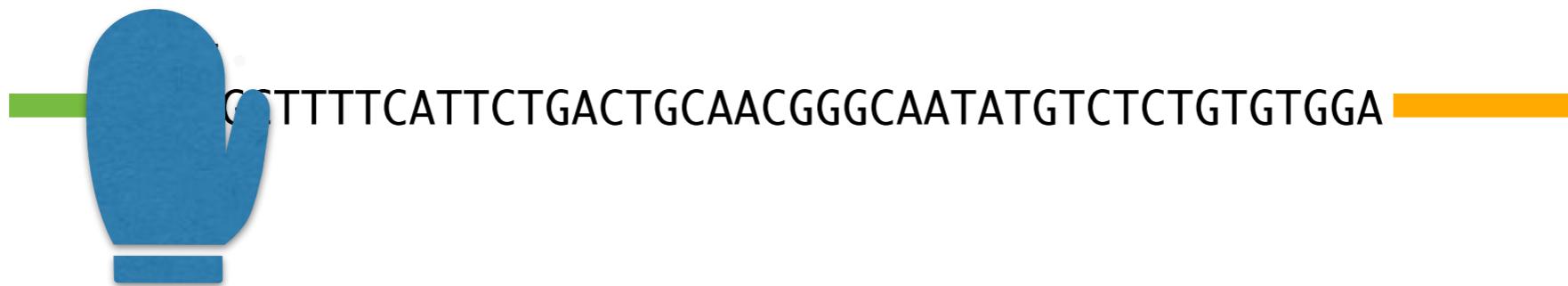
T  
AGCTTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGA

# Sequencing

T  
AGCTTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGA



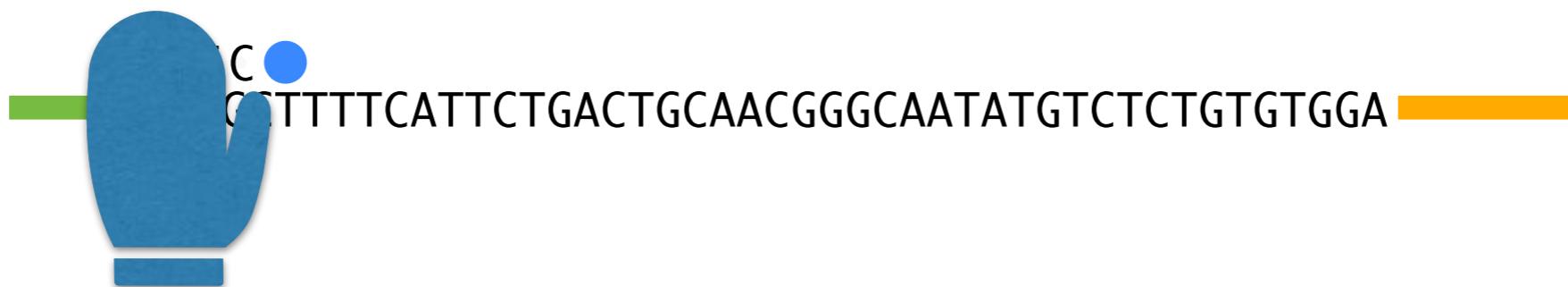
# Sequencing



CCTTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGA

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

# Sequencing



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

# Sequencing



TC [blue dot] AGCTTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGA

# Sequencing



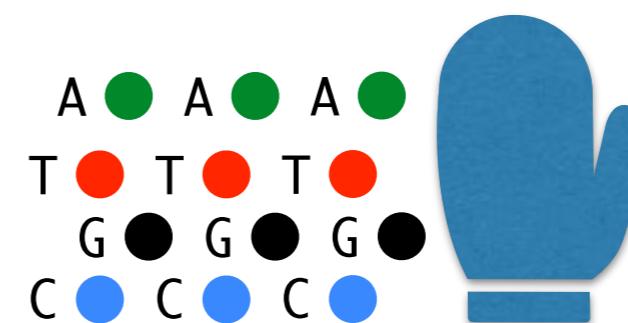
TC  
AGCTTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGA

# Sequencing

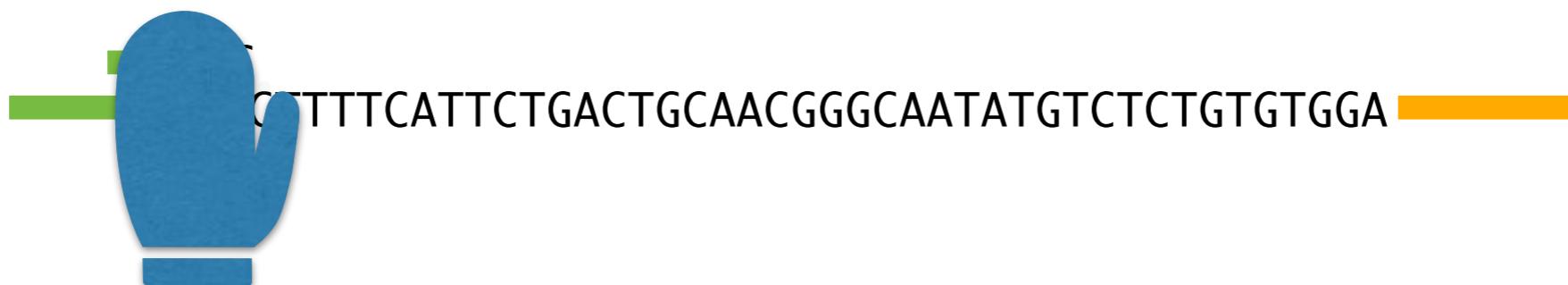
TC  
AGCTTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGA

# Sequencing

TC  
AGCTTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGA

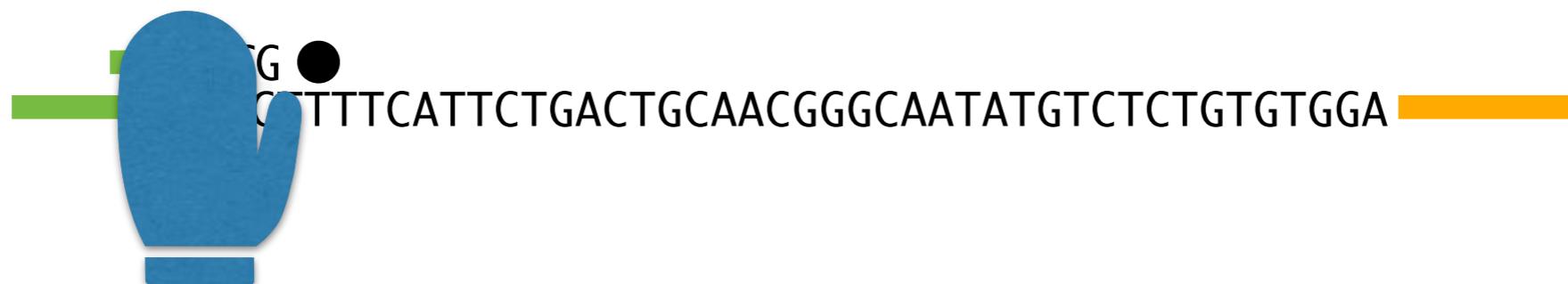


# Sequencing



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

# Sequencing



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

# Sequencing



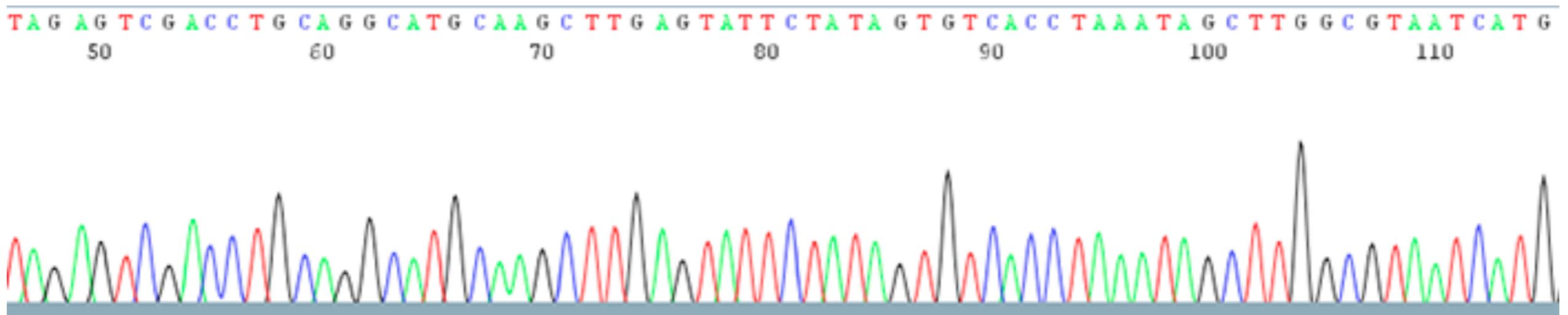
T C G ●  
AGCTTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGA

# Sequencing

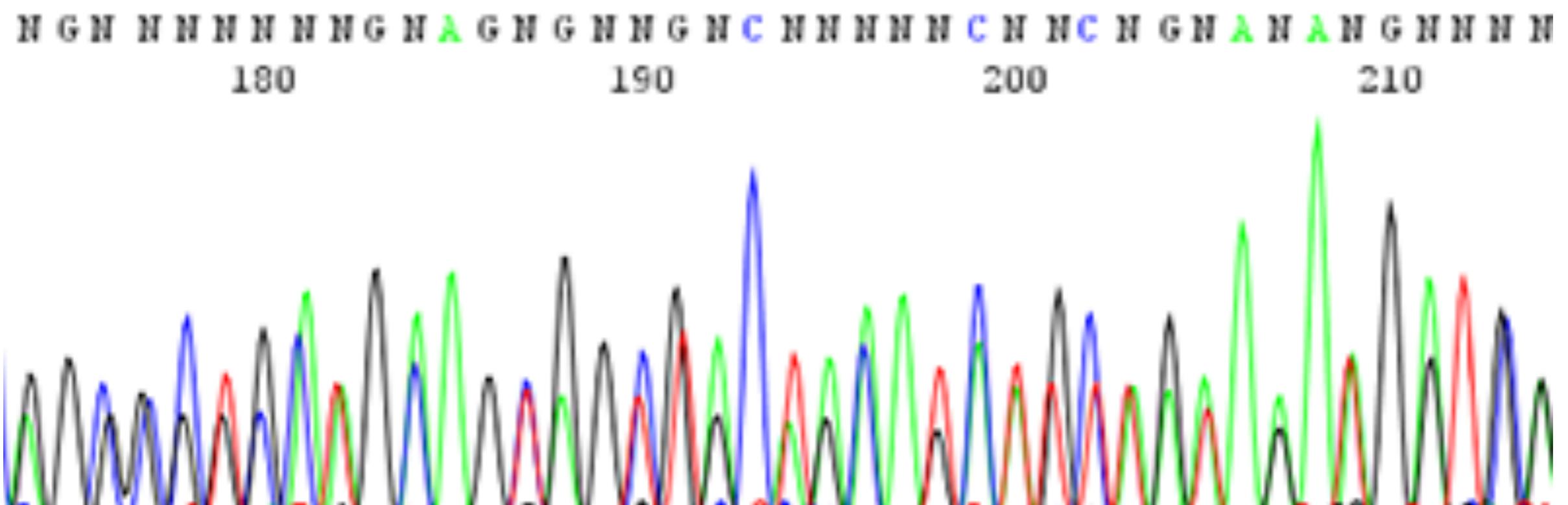


T C G ●  
AGCTTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGA

# Dye-terminator Sanger Sequencing



# Double Sequence



# How?

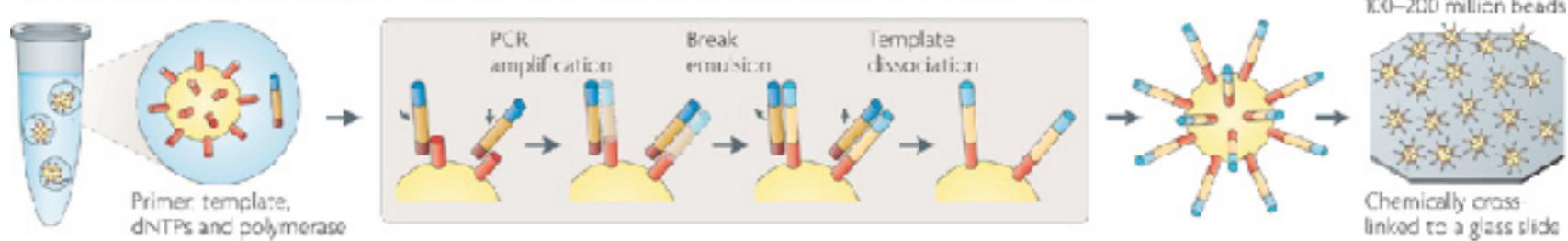
# How?

- Separate
- Detect
- Removable Terminator

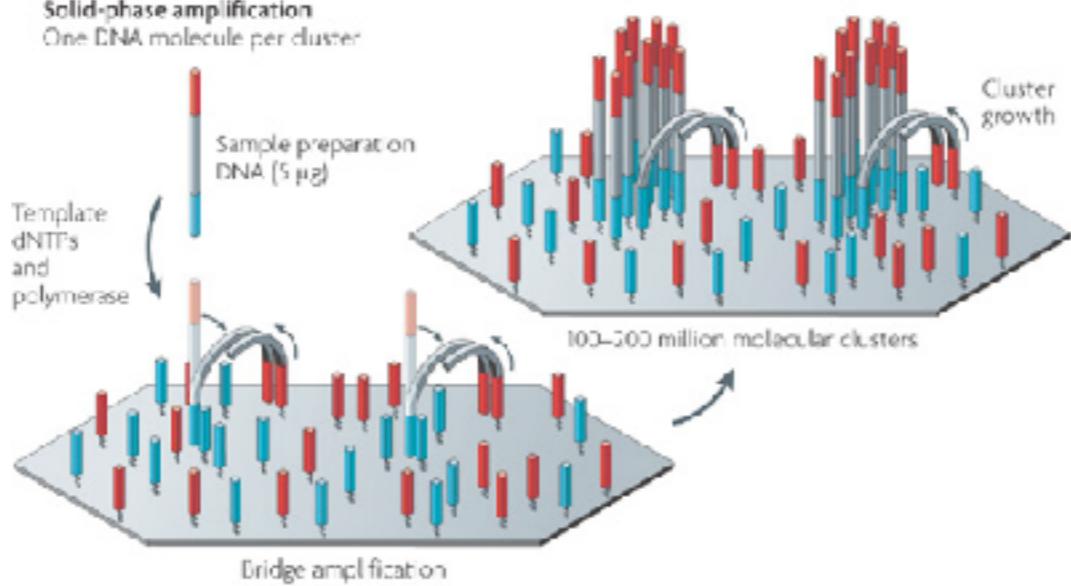
# Template immobilization

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

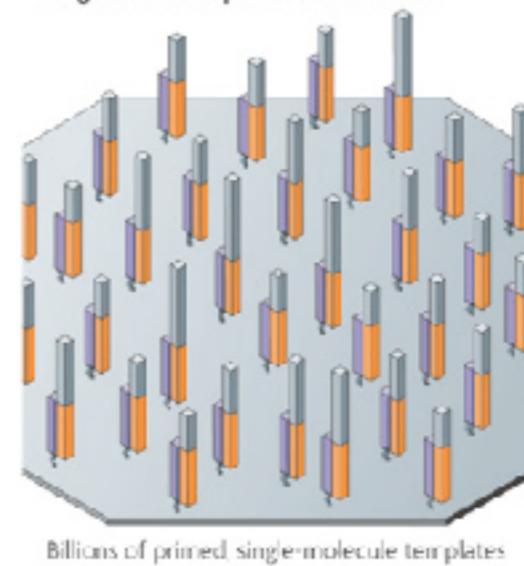
One DNA molecule per bead. Clonal amplification to thousands of copies occurs in microreactors in an emulsion



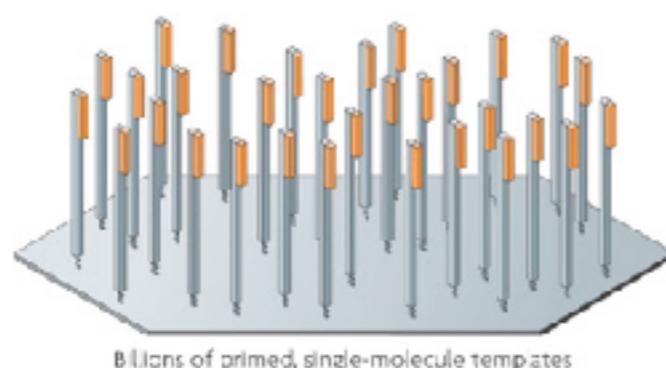
b Illumina/Solexa  
Solid-phase amplification  
One DNA molecule per cluster



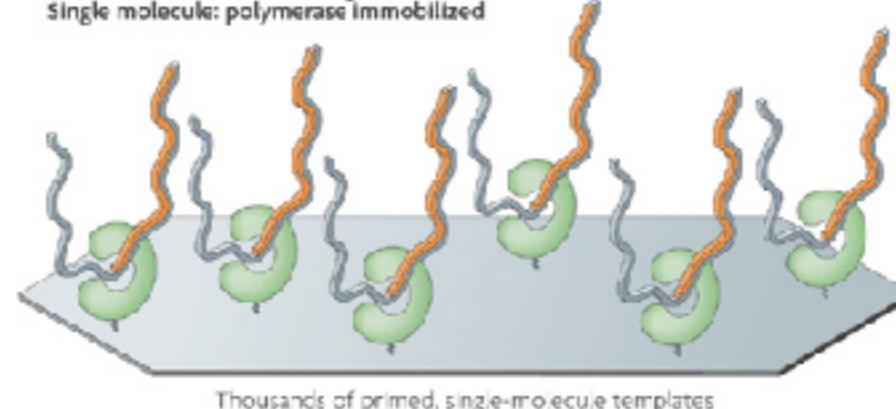
c Helicos BioSciences: one-pass sequencing  
Single molecule: primer immobilized



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



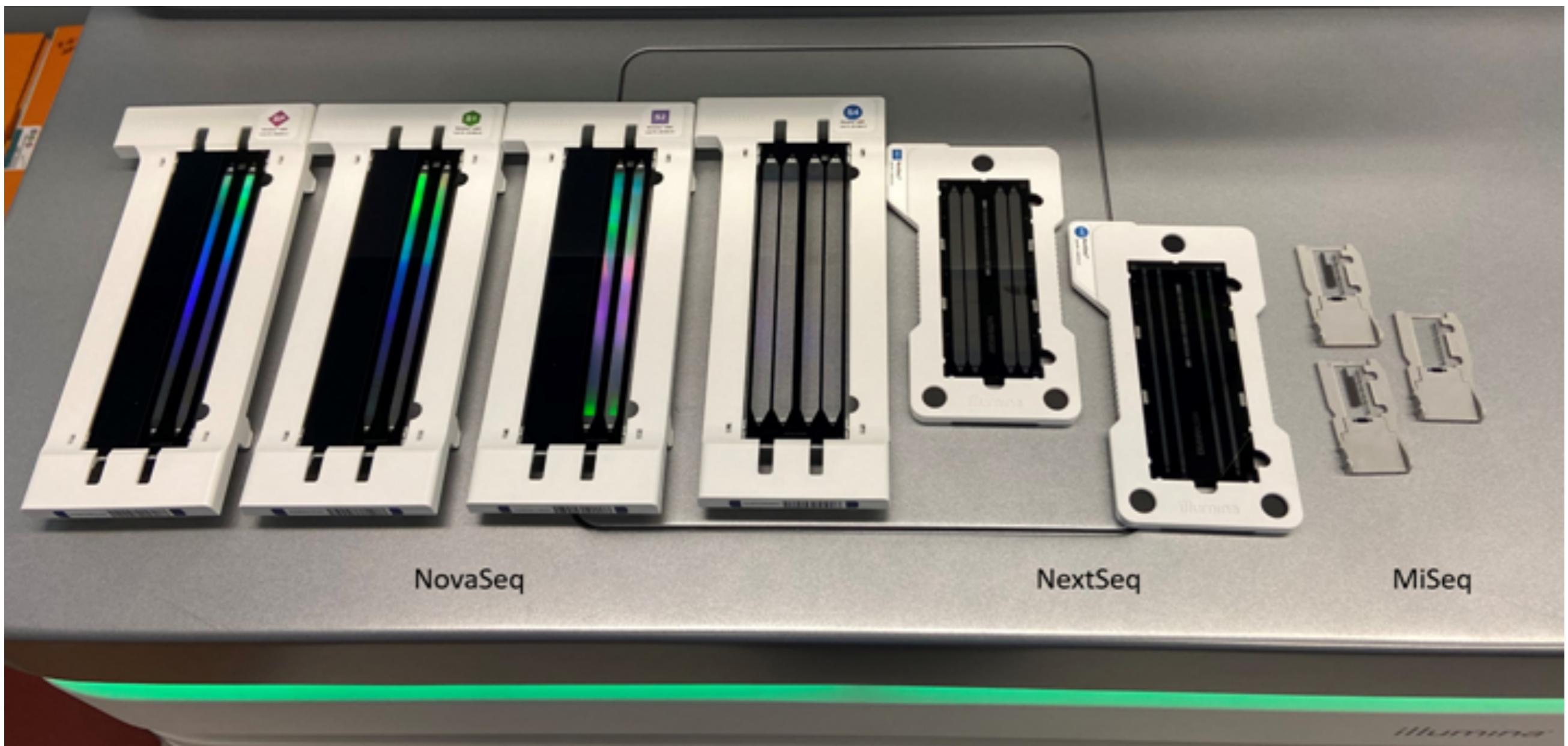
# A Flow Cell



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

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

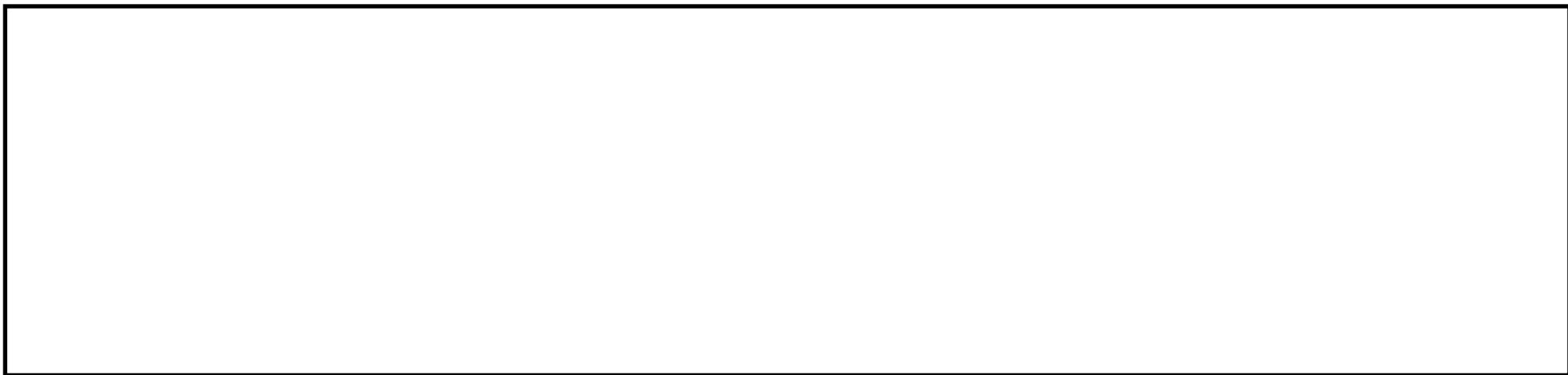
# A Flow Cell



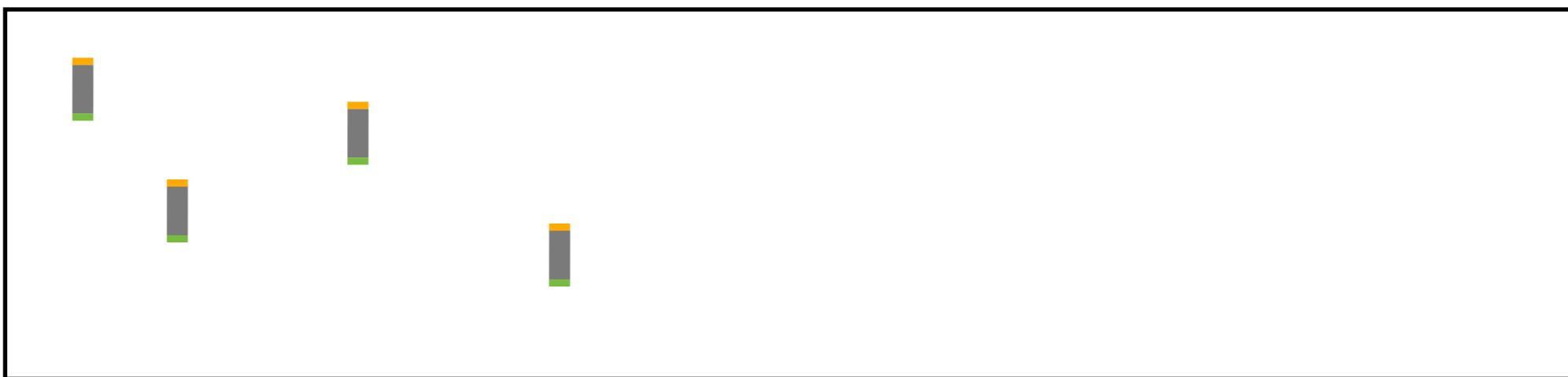
# SBS: Sequencing by Synthesis

An Illumina Story

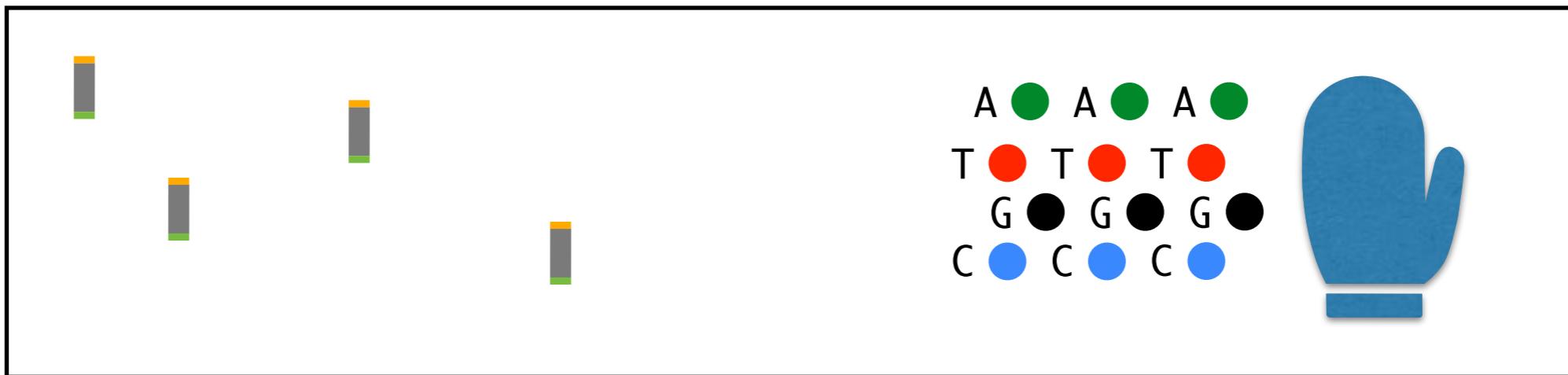
# A Flow Cell



# Bind Library



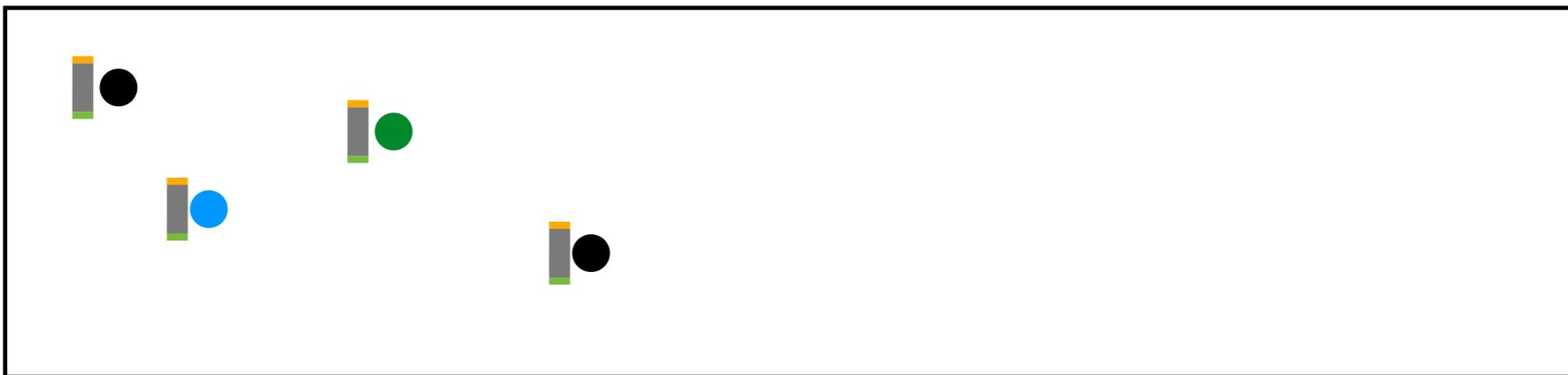
# 1st Cycle



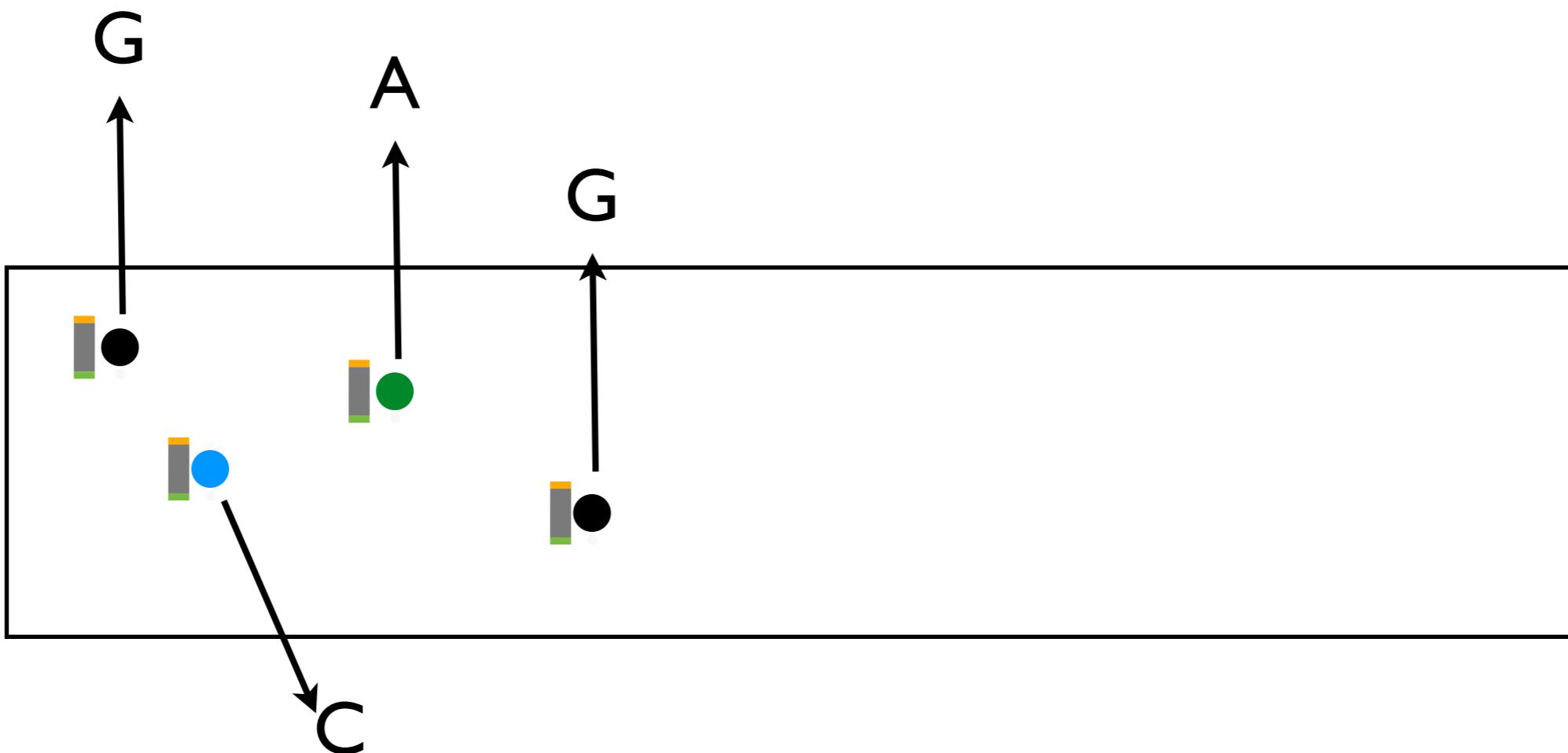
# 1st Cycle



# 1st Cycle



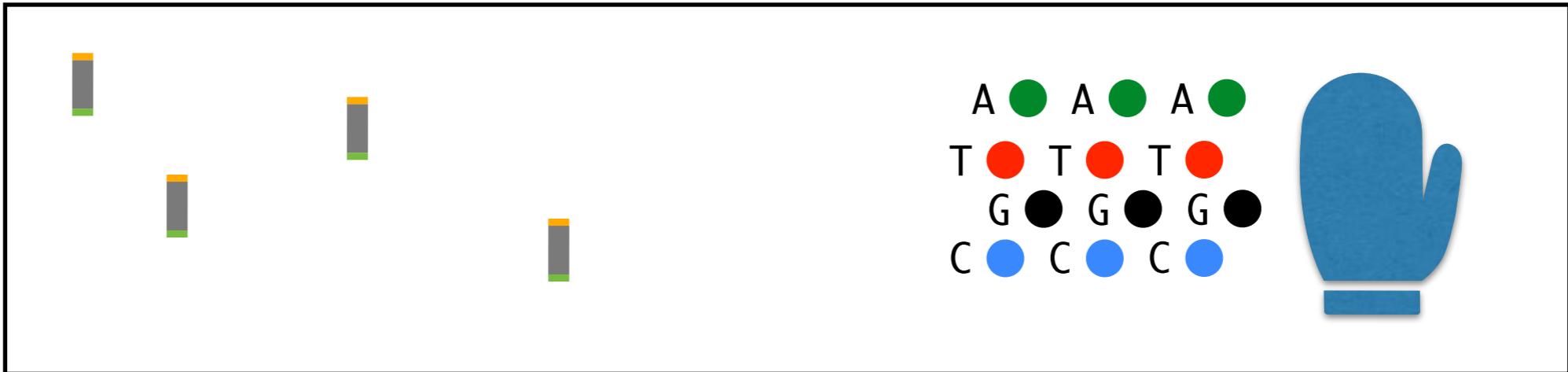
# 1st Cycle



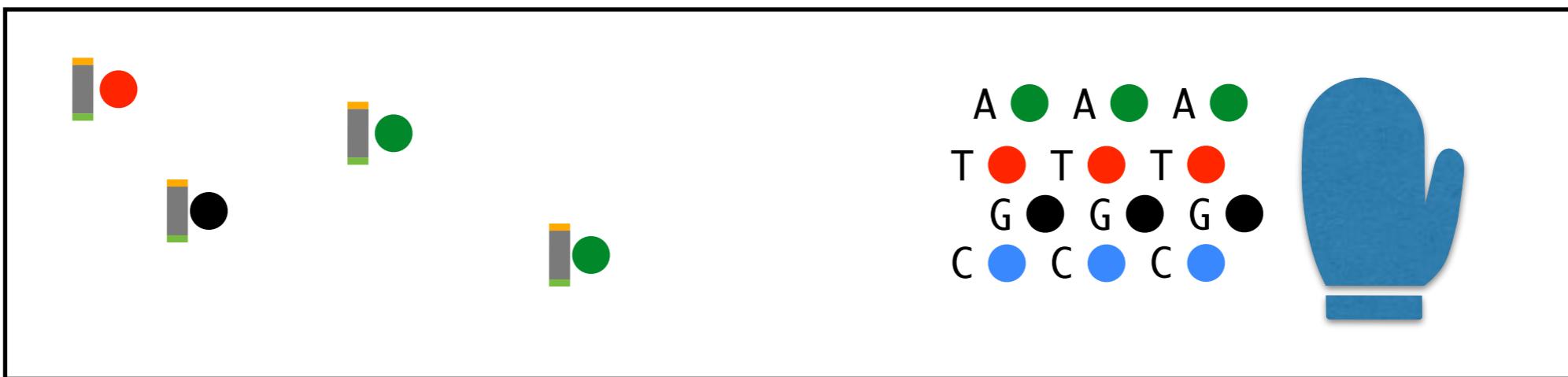
# 2nd Cycle



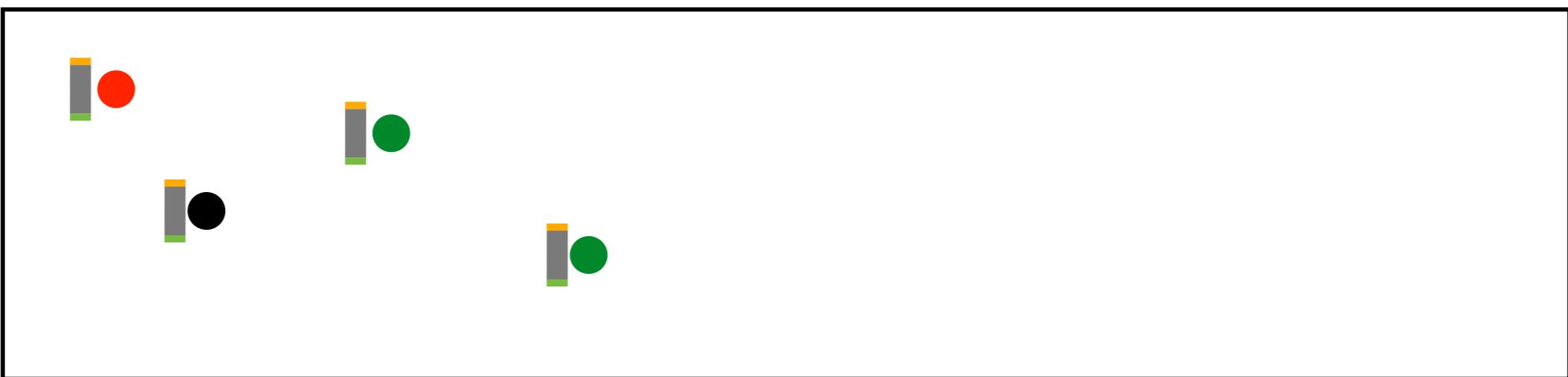
# 2nd Cycle



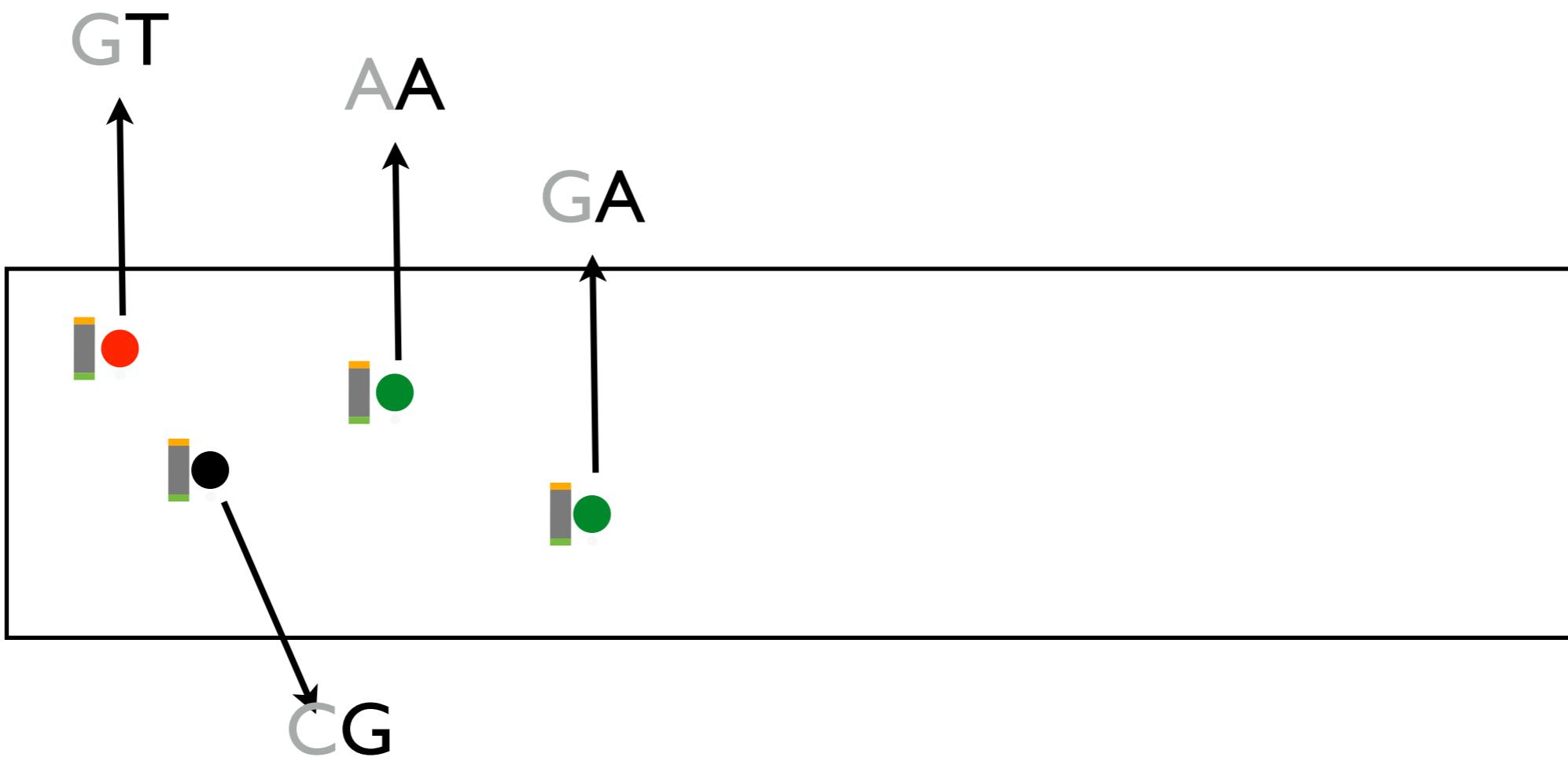
# 2nd Cycle



# 2nd Cycle



# 2nd Cycle



# 3rd Cycle



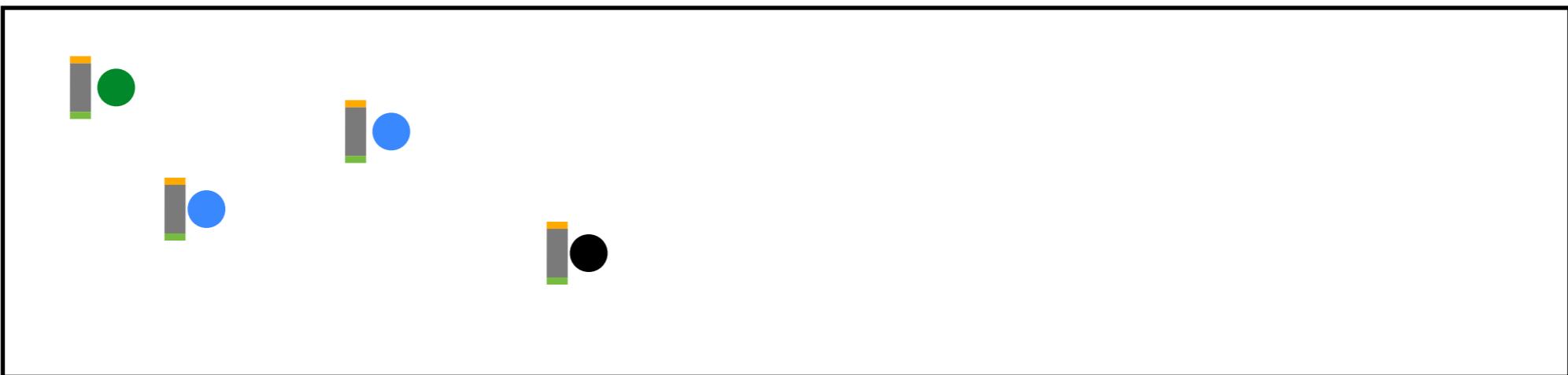
# 3rd Cycle



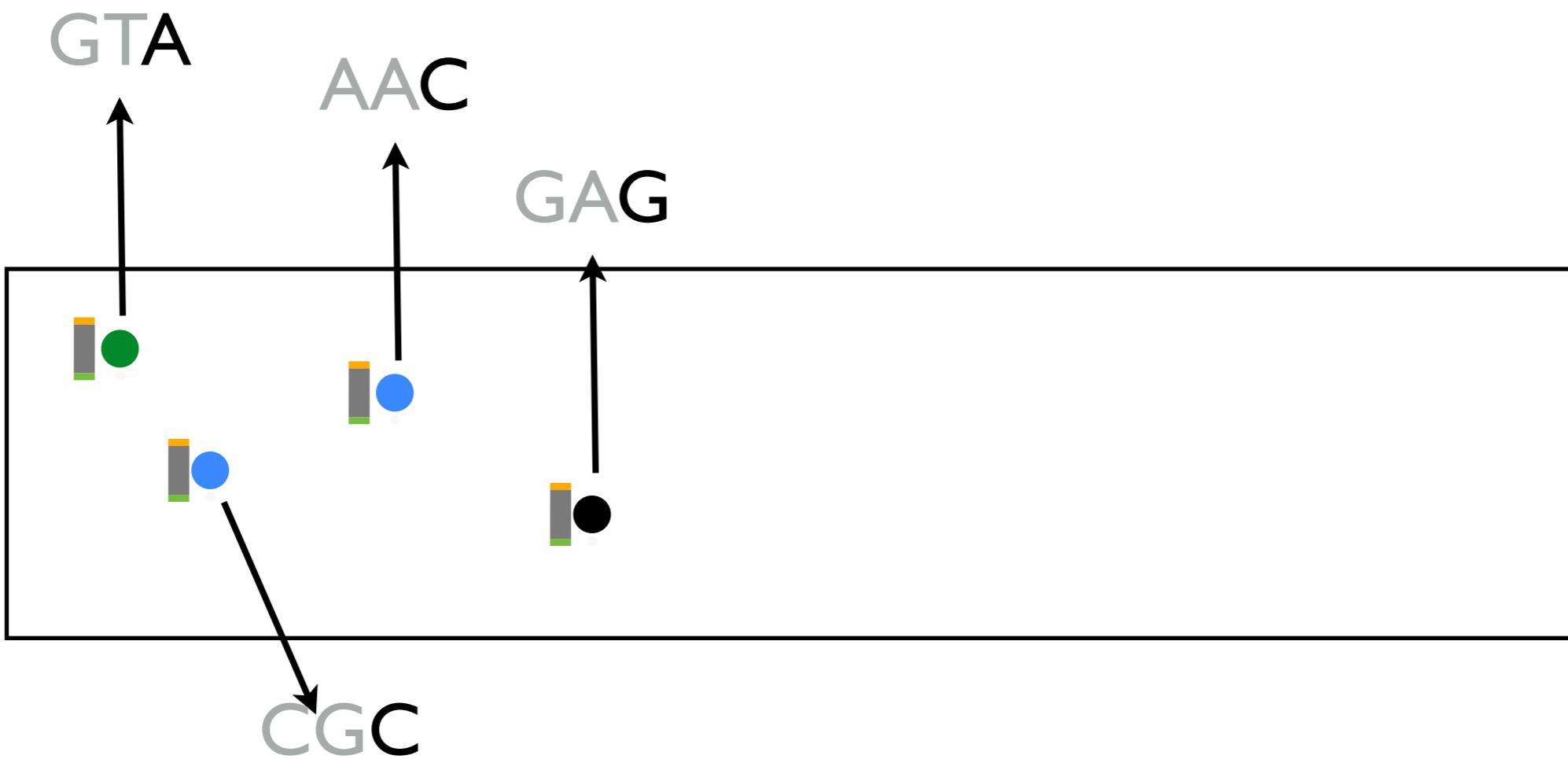
# 3rd Cycle



# 3rd Cycle



# 3rd Cycle

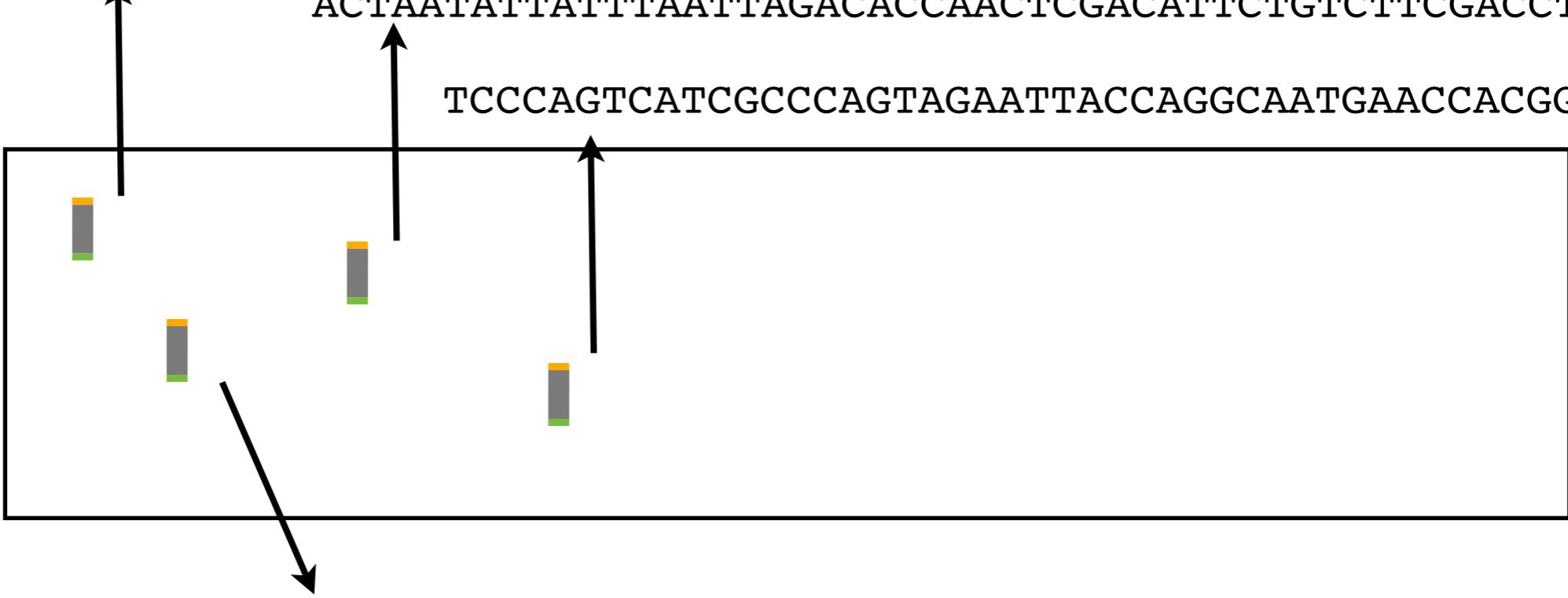


# 50th Cycle

GAATTCTAAAACAGTTGCATTCTATAATTACAAAATAATTGAAACACTTC

ACTAATATTATTAAATTAGACACCAACTCGACATTCTGTCTTCGACCTAT

TCCCAGTCATCGCCCCAGTAGAATTACCAGGCAATGAACCACGGCCTTCA

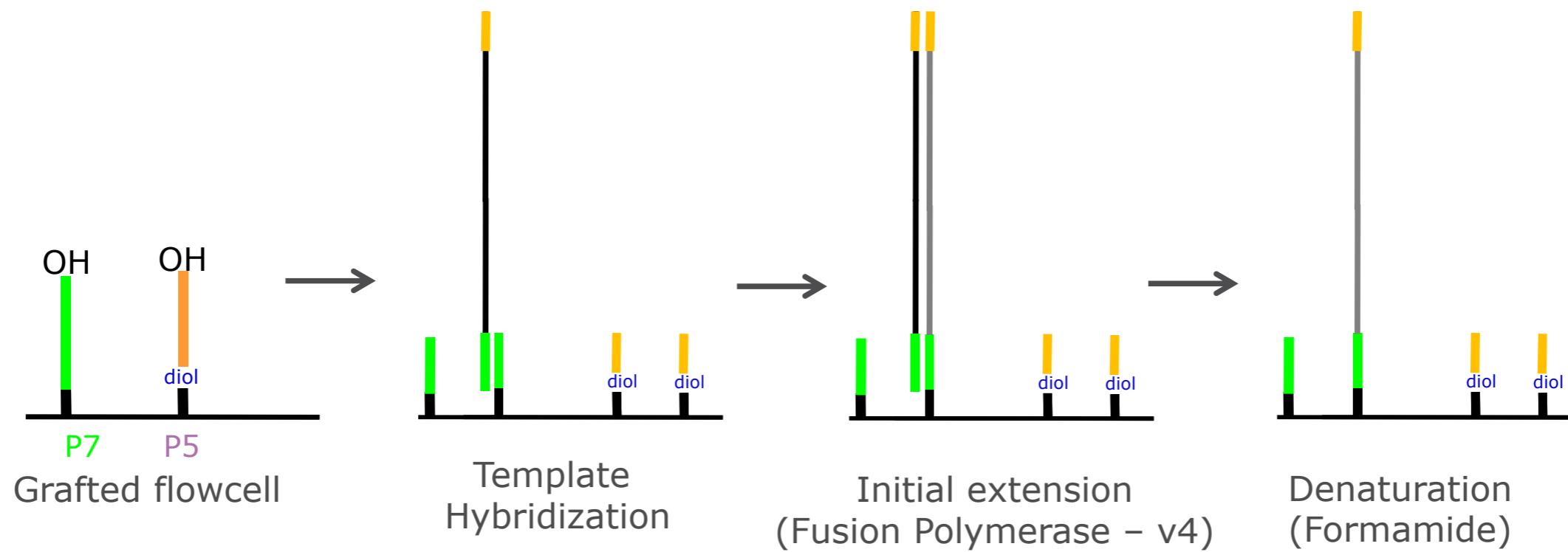


ACAGCATATGGGTTCACTCCAACAGTGAACCATTCCAAAAGGCCTTGCCT

# Illumina Short Reads

- 50 - 300bp

# Cluster generation – hybridization and amplification



# Hybridization

5' -CTGATCTGACTGATGCGTATGCTAGT-3'

+

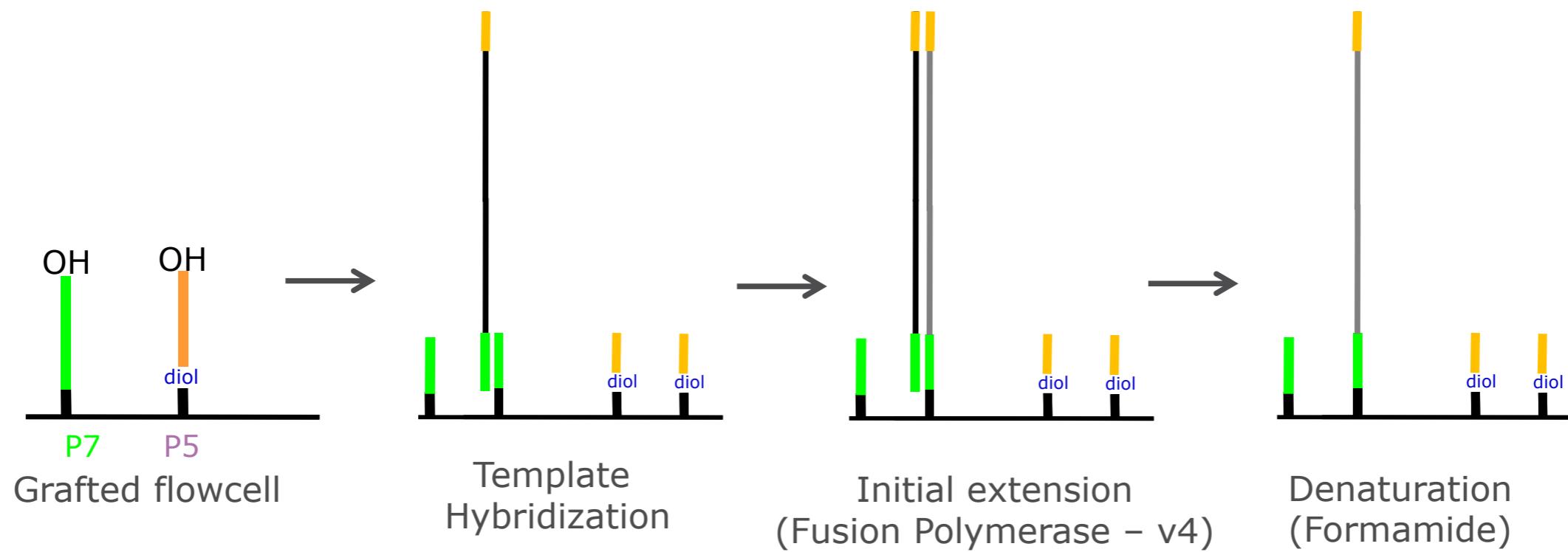
3' -**GCATAC**-5'

=

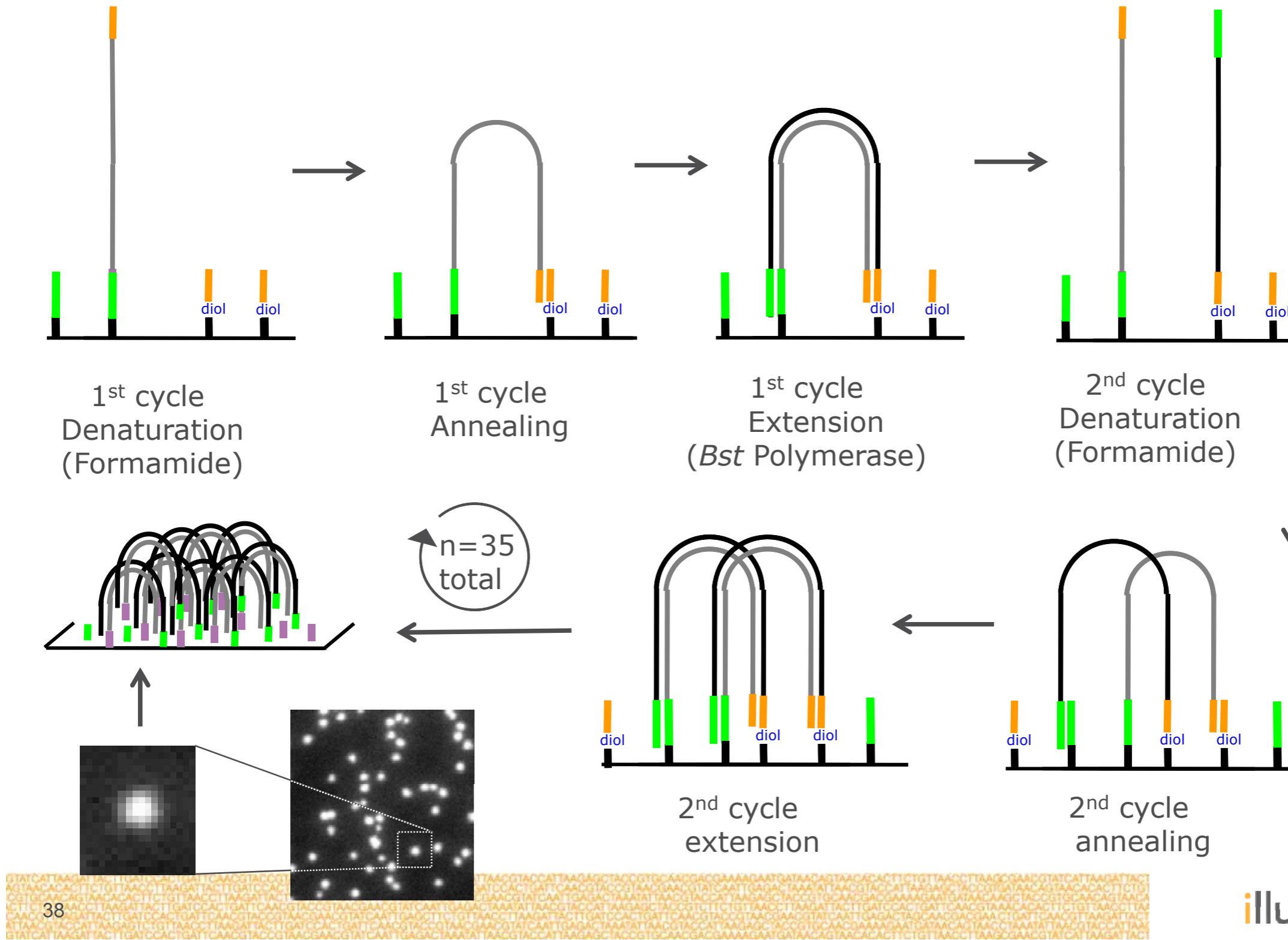
5' -CTGATCTGACTGATGCGTATGCTAGT-3'

3' -**GCATAC**-5'

# Cluster generation – hybridization and amplification



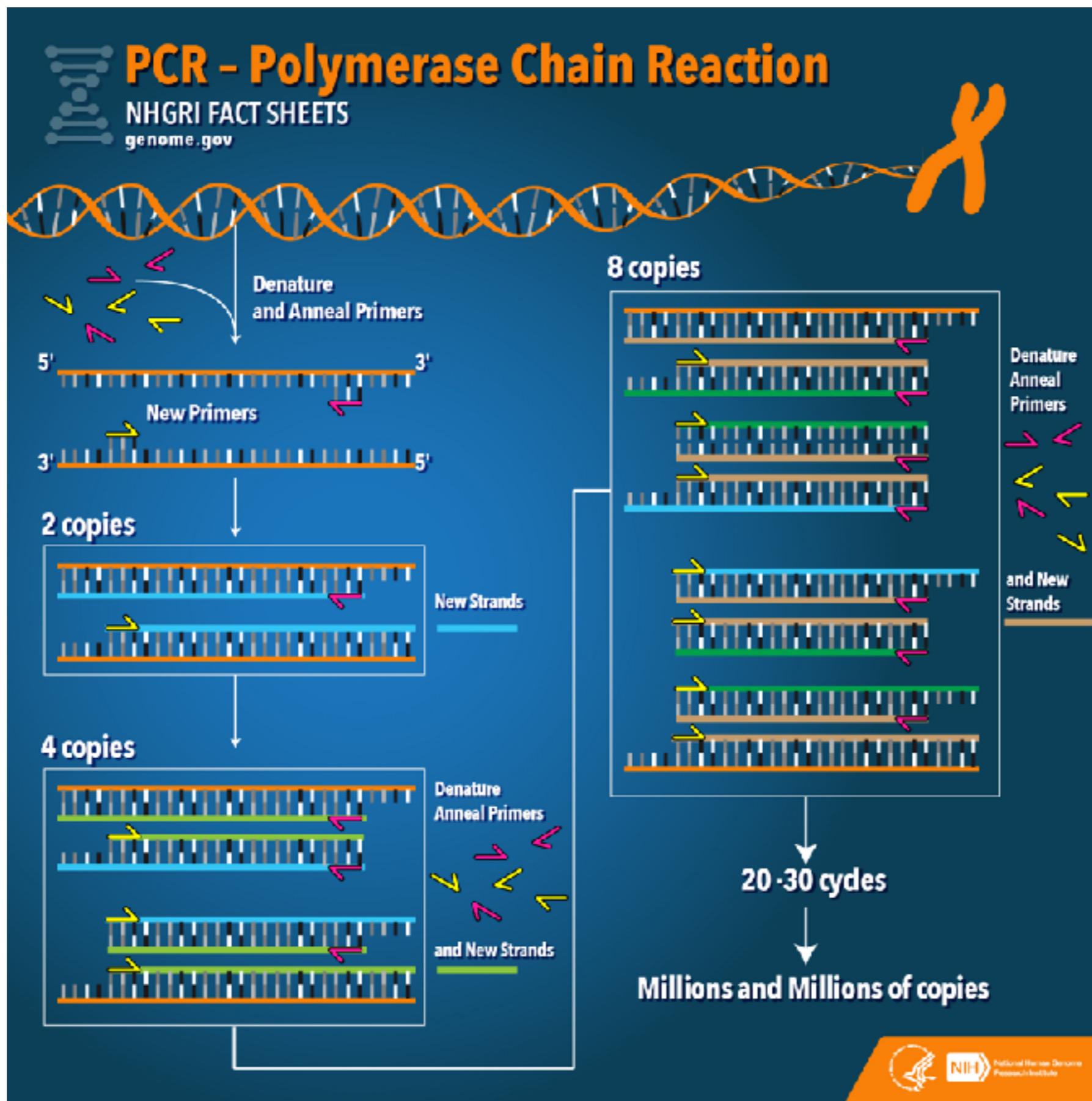
# Cluster generation – hybridization and amplification



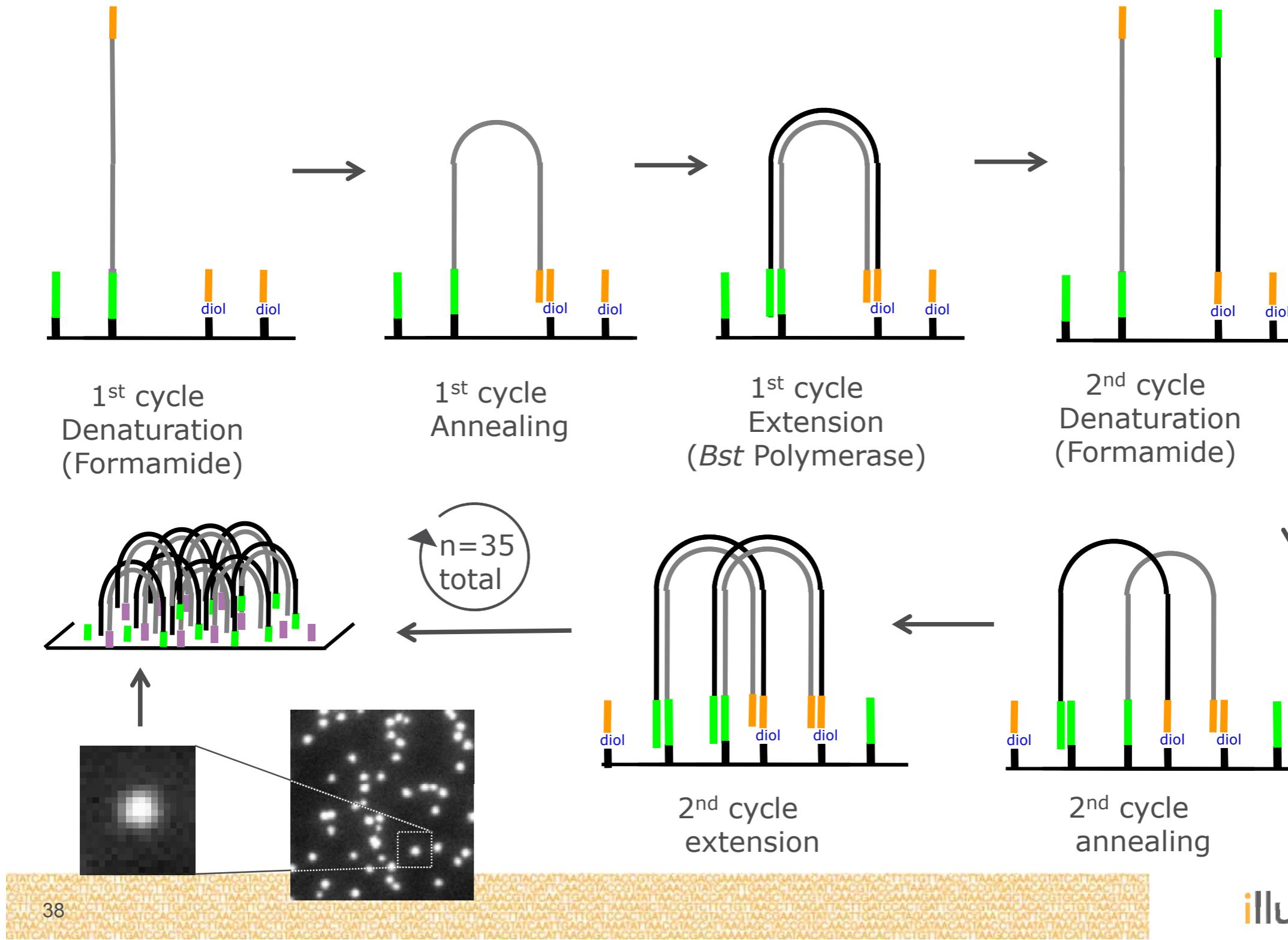
38

illumina®

# PCR



# Cluster generation – hybridization and amplification



38

illumina®

# Library Preparation

## Purified Nucleic Acid



Adapter Ligation



# Why Adapters?

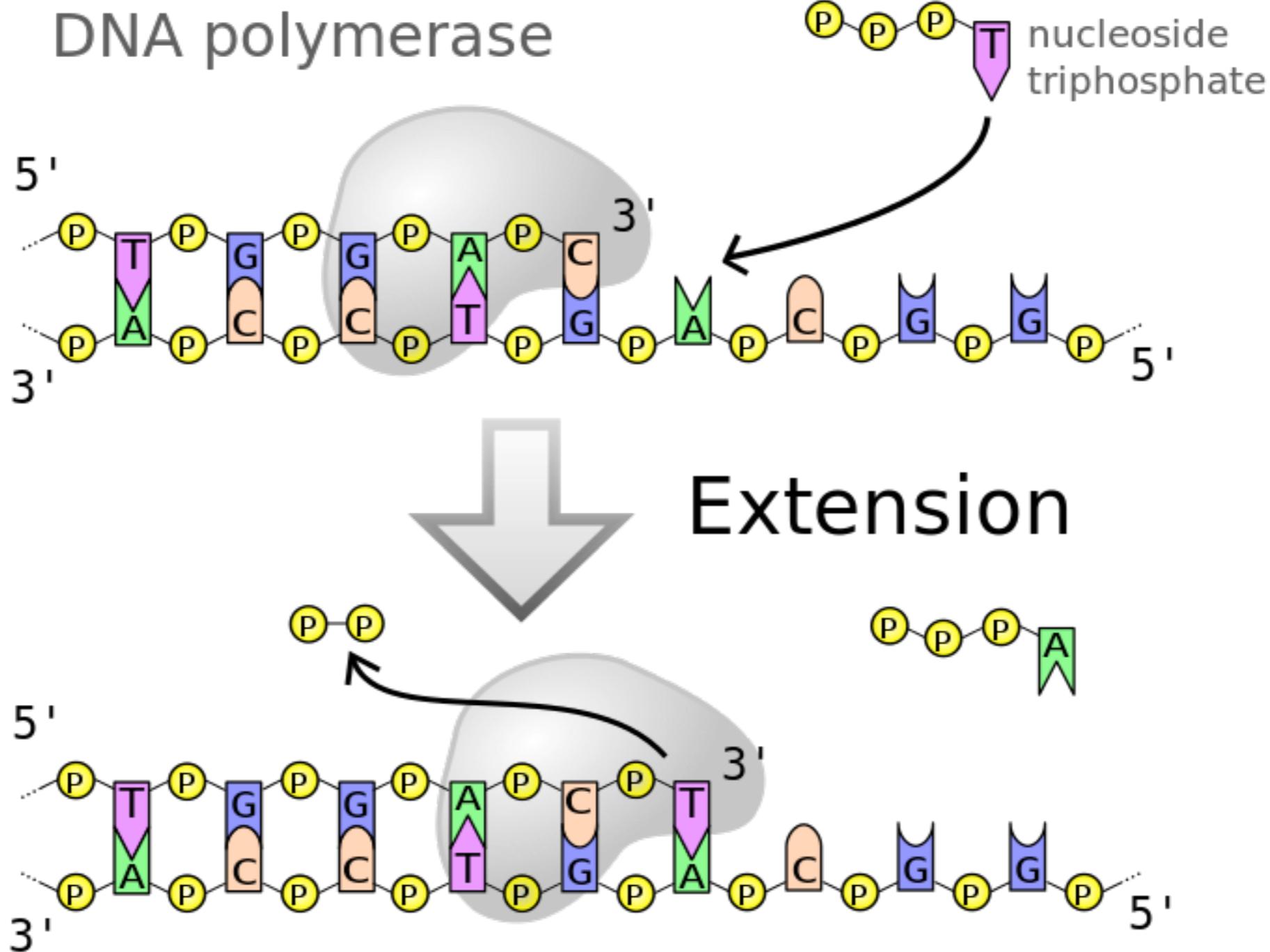
# DNA Synthesis

- 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



# Why Adapters?

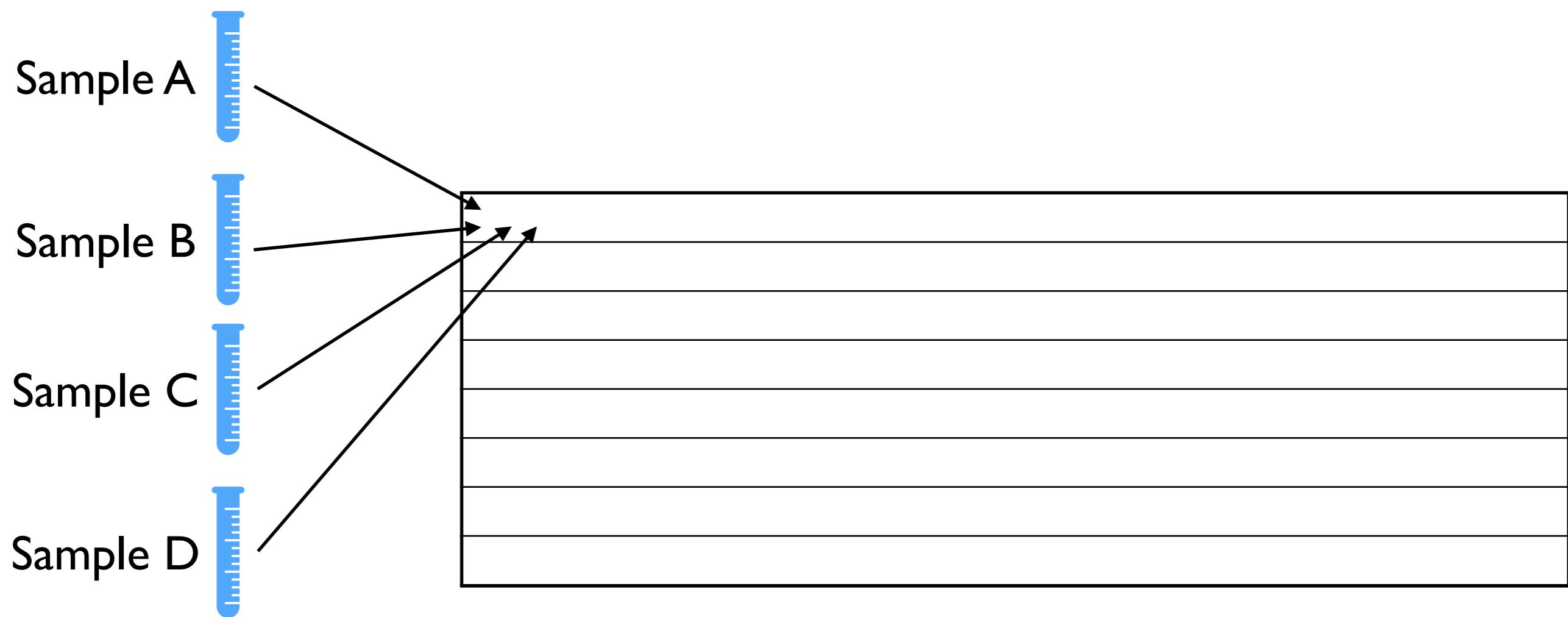
- Universal Priming Sites
  - Sequencing Primers
  - PCR Primers
- Hybridization to Flow Cell
- Index Barcodes

# Why Adapters?

- Universal Priming Sites
  - Sequencing Primers
  - PCR Primers
- Hybridization to Flow Cell
- Index Barcodes

# Barcodes: Why?

Multiplexing: Combine multiple samples in a lane

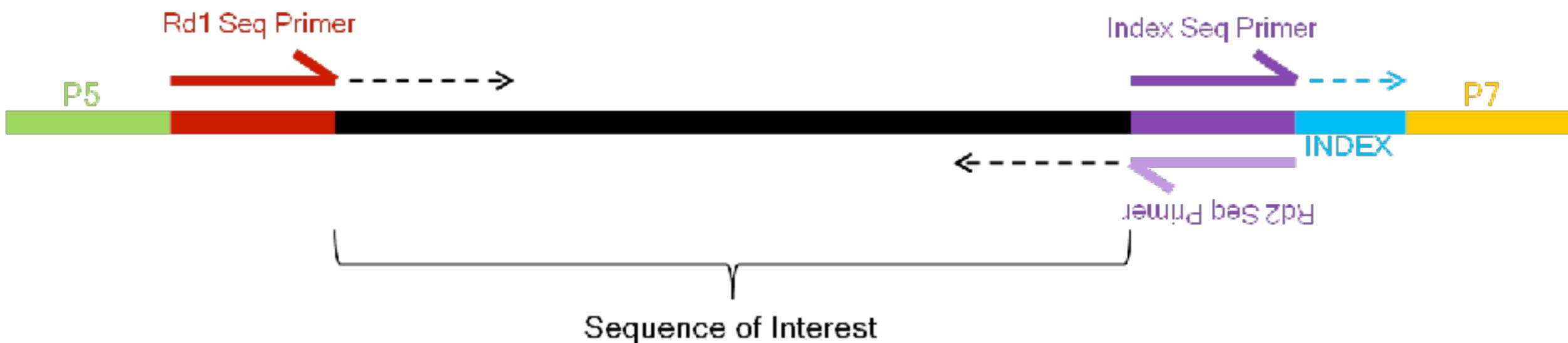


# Barcodes

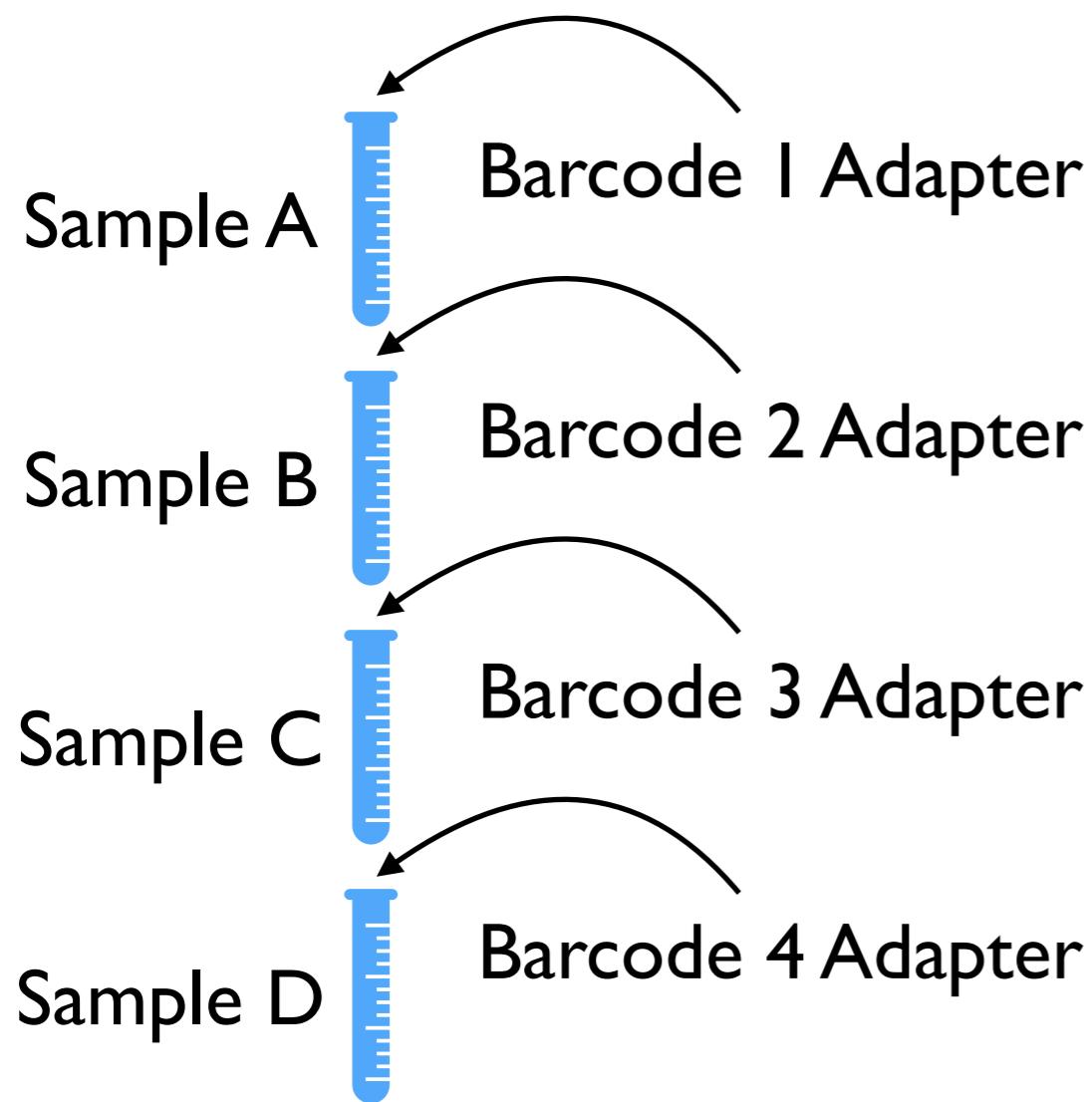
Sample_Name	index
control_1	AAGACCGT
control_2	TTGCGAGA
control_3	GCAATTCC
treatmentA_1	GAATCCGT
treatmentA_2	CCGCTTAA
treatmentA_3	TACCTGCA
treatmentB_1	GTCGATTG
treatmentB_2	TATGGCAC
treatmentB_3	CTCGAACCA

# Multiplexing (Barcodes)

## STRUCTURE DETAILS



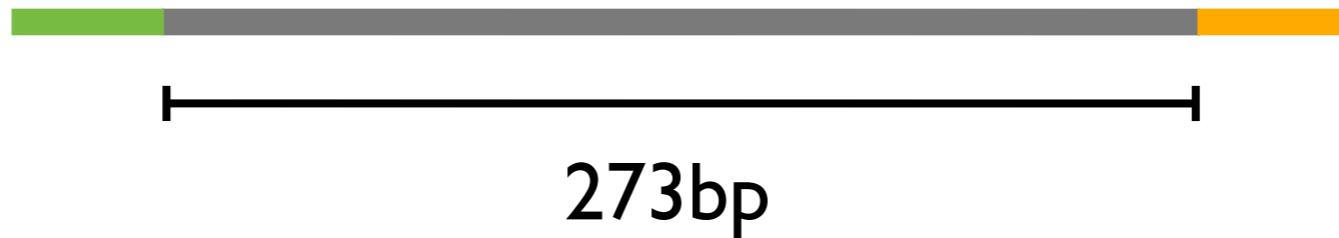
# Barcodes



# **Additional Sequencing Details**

# Read Length

bases  
50 →



# Read Length

bases

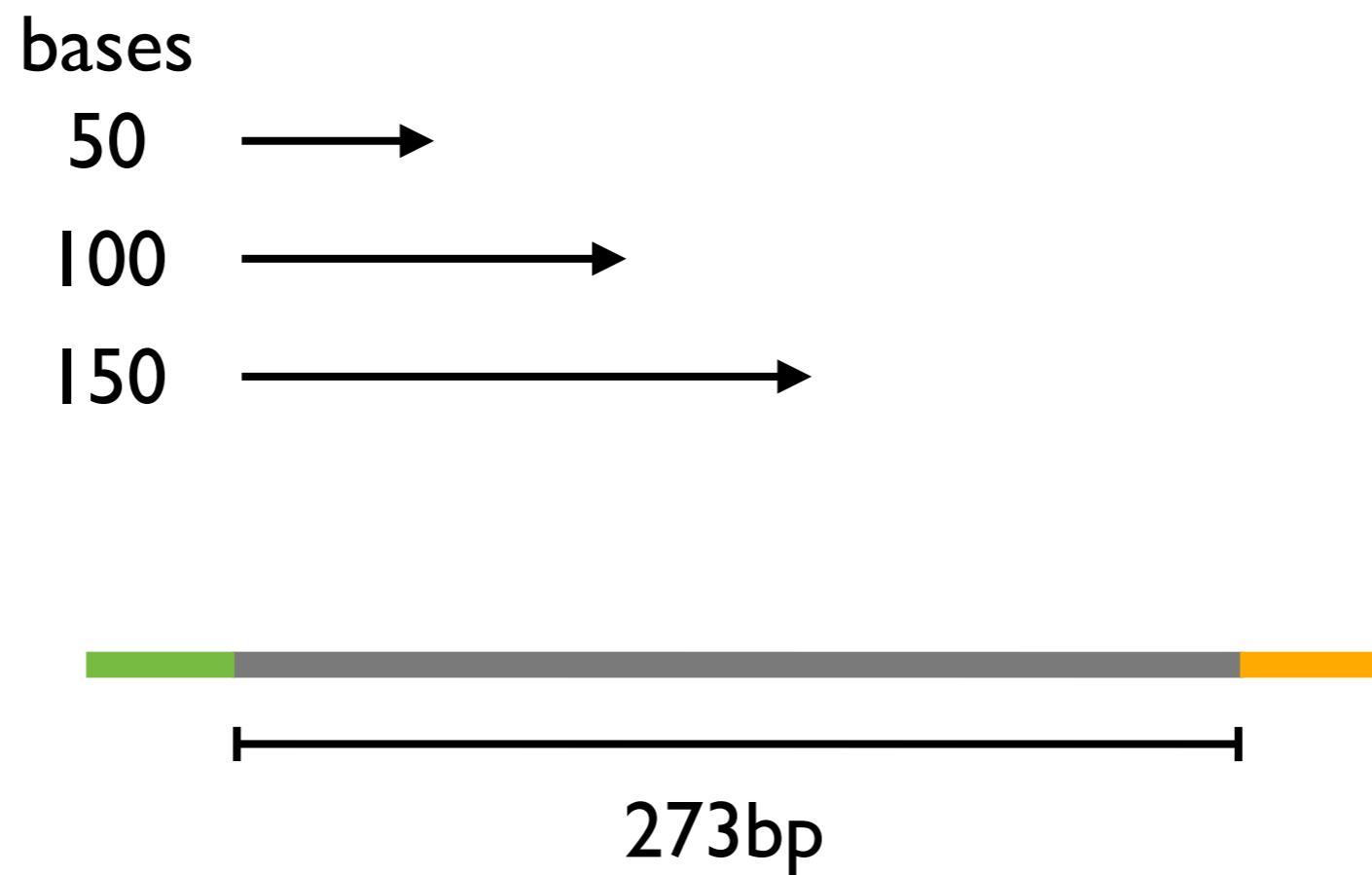
50 →

100 →

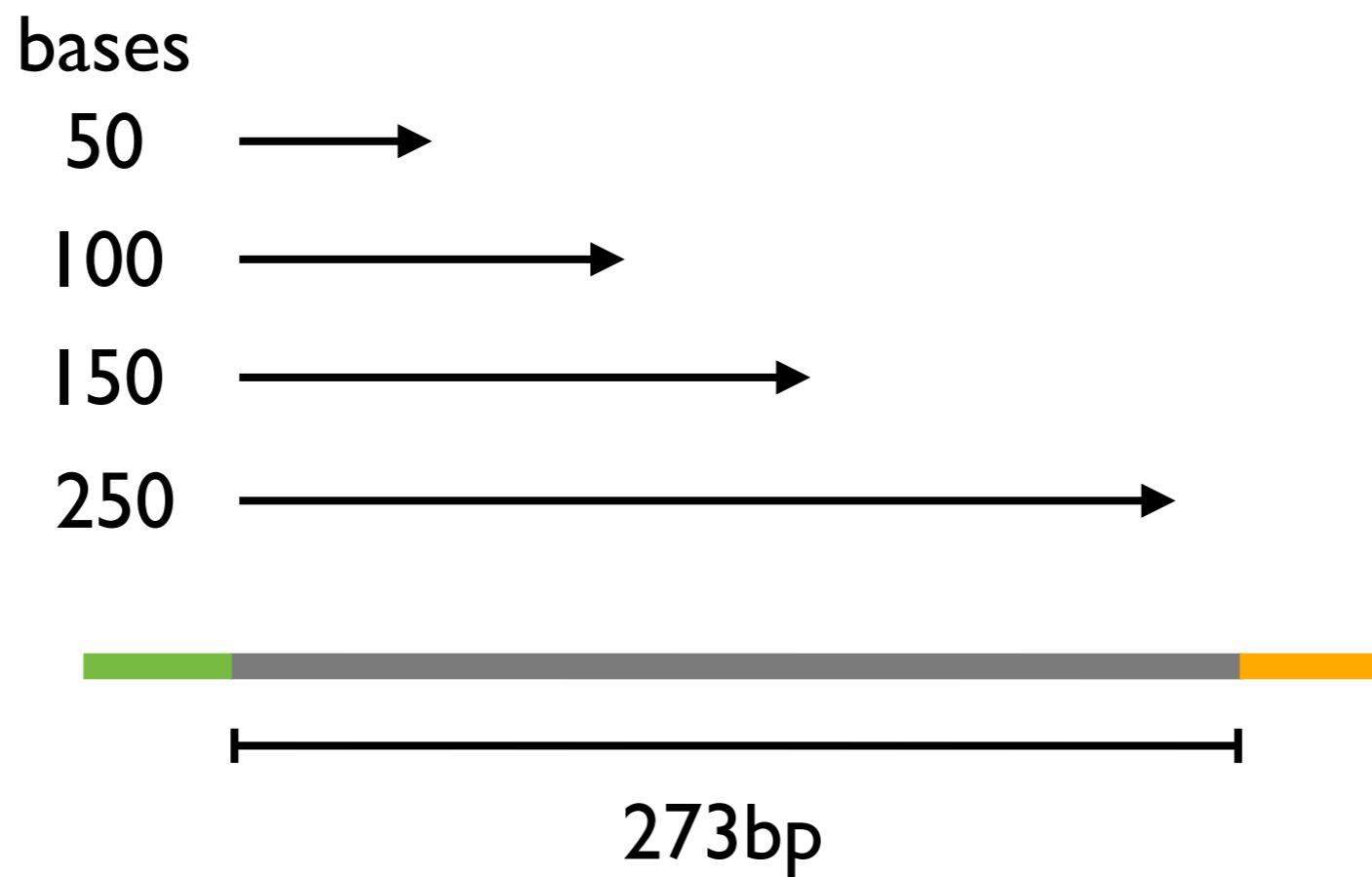


273bp

# Read Length



# Read Length



# Paired-End

TCGAAAAG  
AGCTTTCTTGTACTGCAACGGGCAATATGTCTGTGTGGA

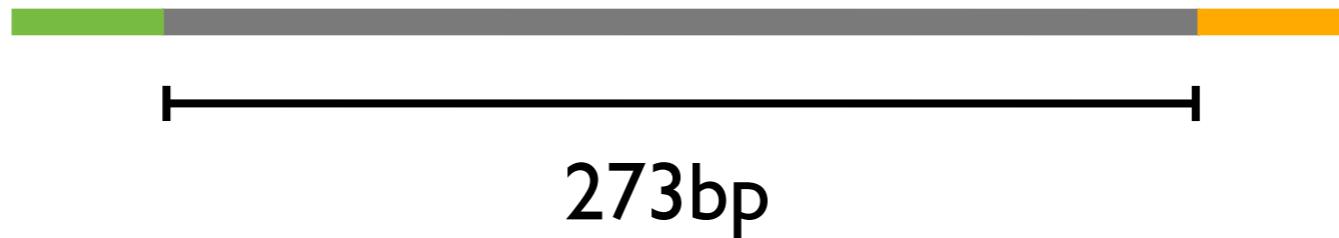
# Paired-End

TCGAAAAG  
AGCTTTCAATTCTGACTGCAACGGGCAATATGTCTGTGTGGA

AGCTTTCAATTCTGACTGCAACGGGCAATATGTCTGTGTGGA  
GACACACCT

# Read Length

bases  
50 →



# Read Length

bases

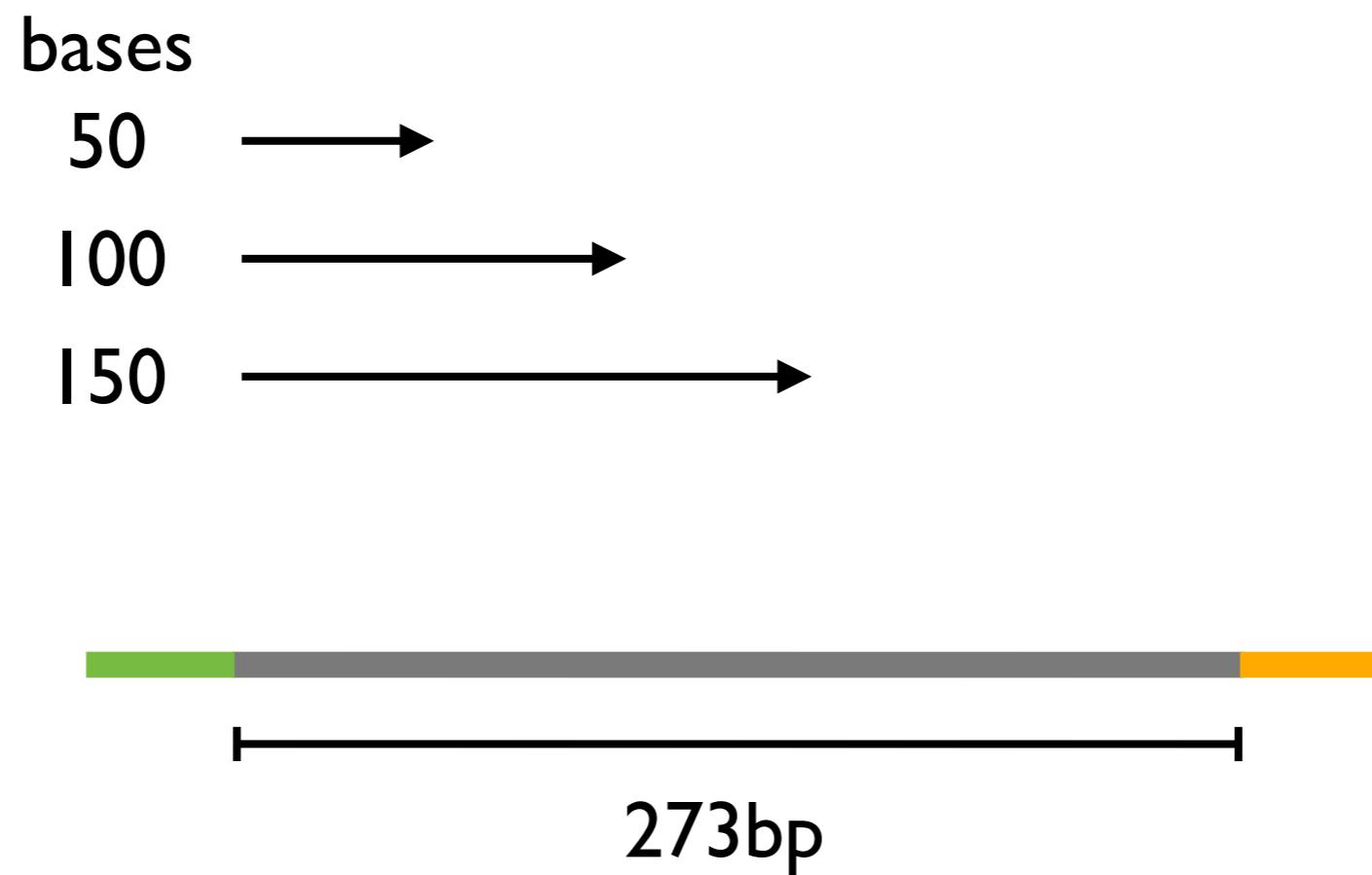
50 →

100 →

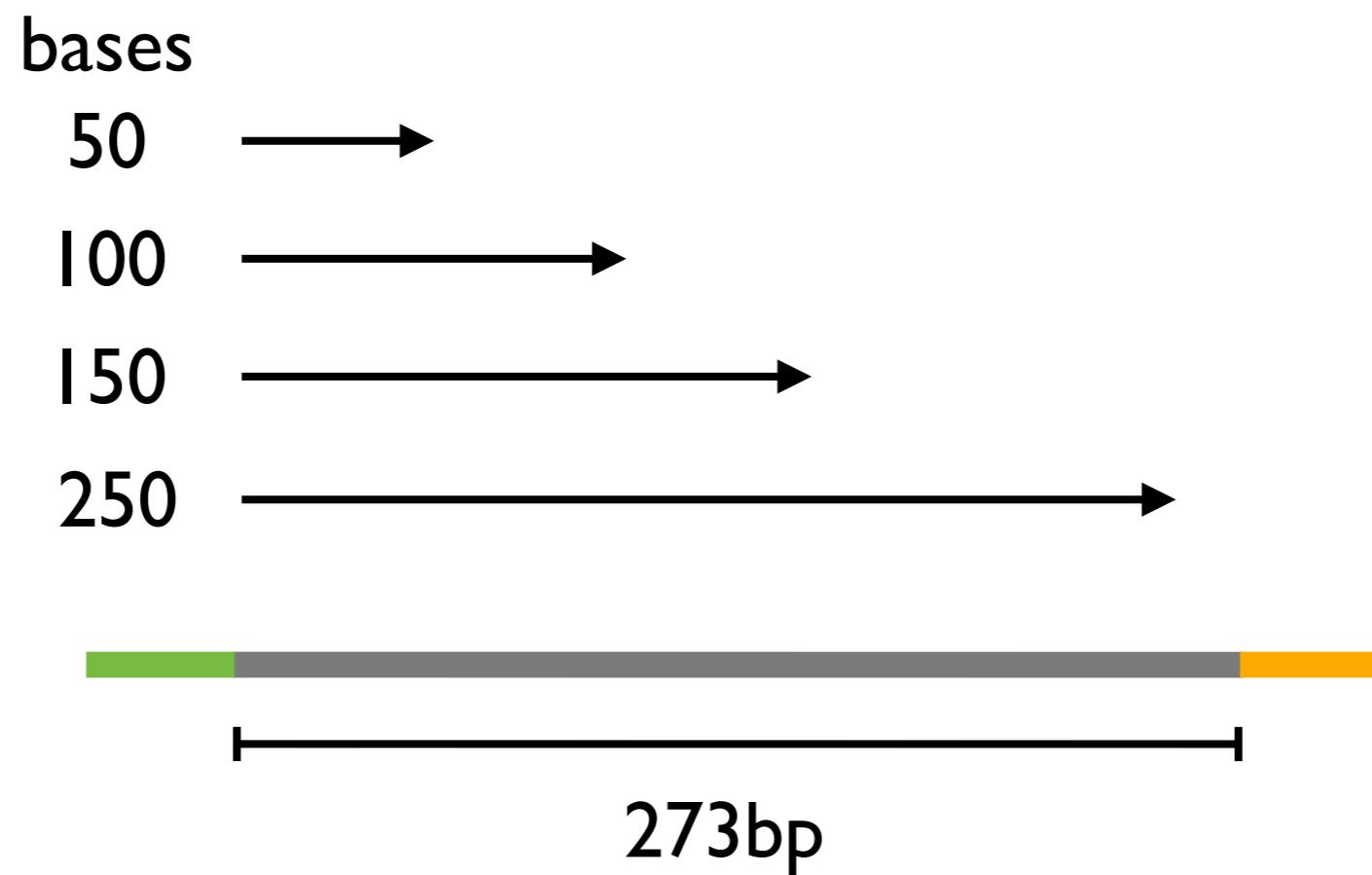


273bp

# Read Length



# Read Length



# Read Length

bases



# Read Length

bases  
50 →



# Read Length

bases

50 →

100 →



# Read Length

bases

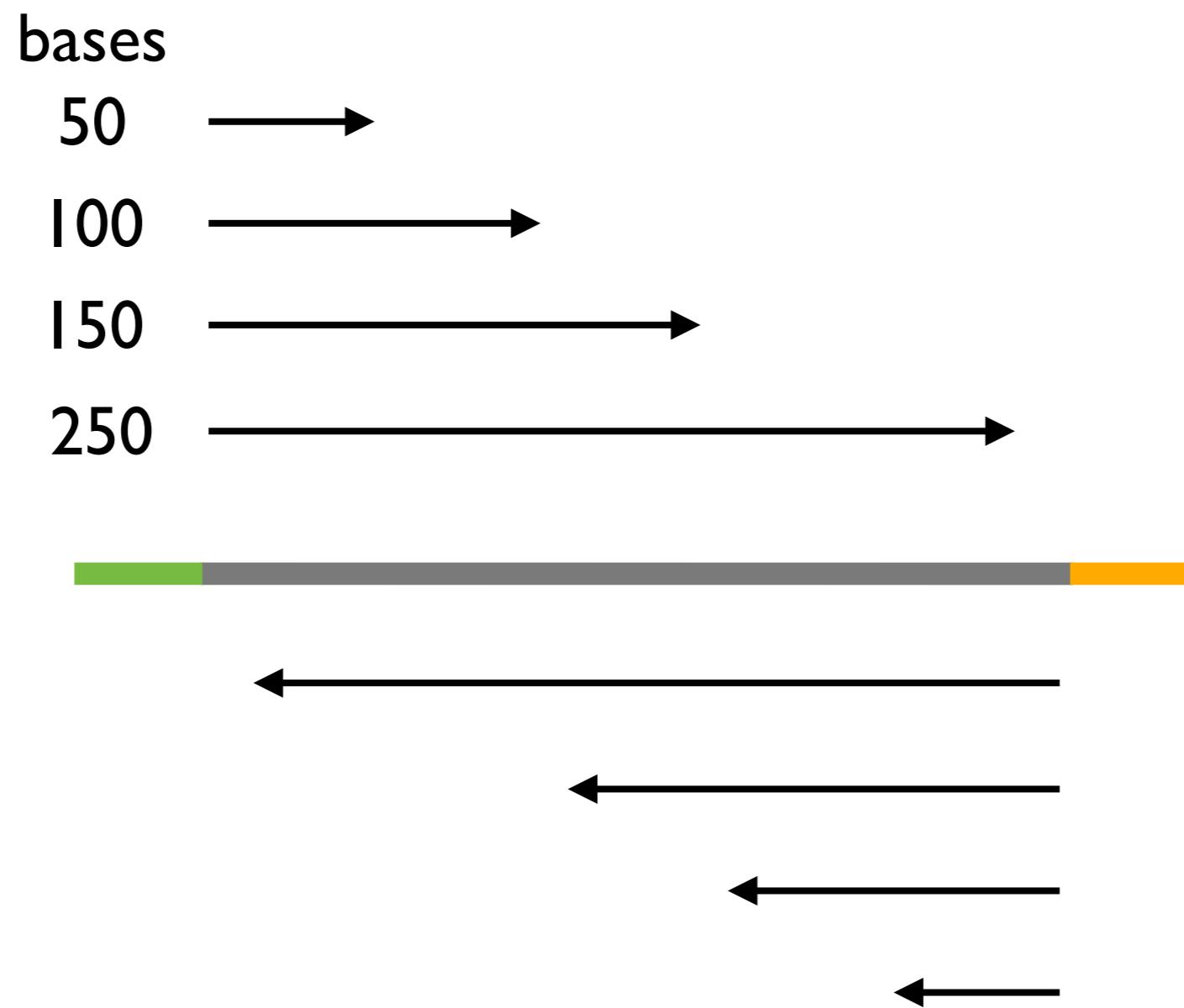
50 →

100 →

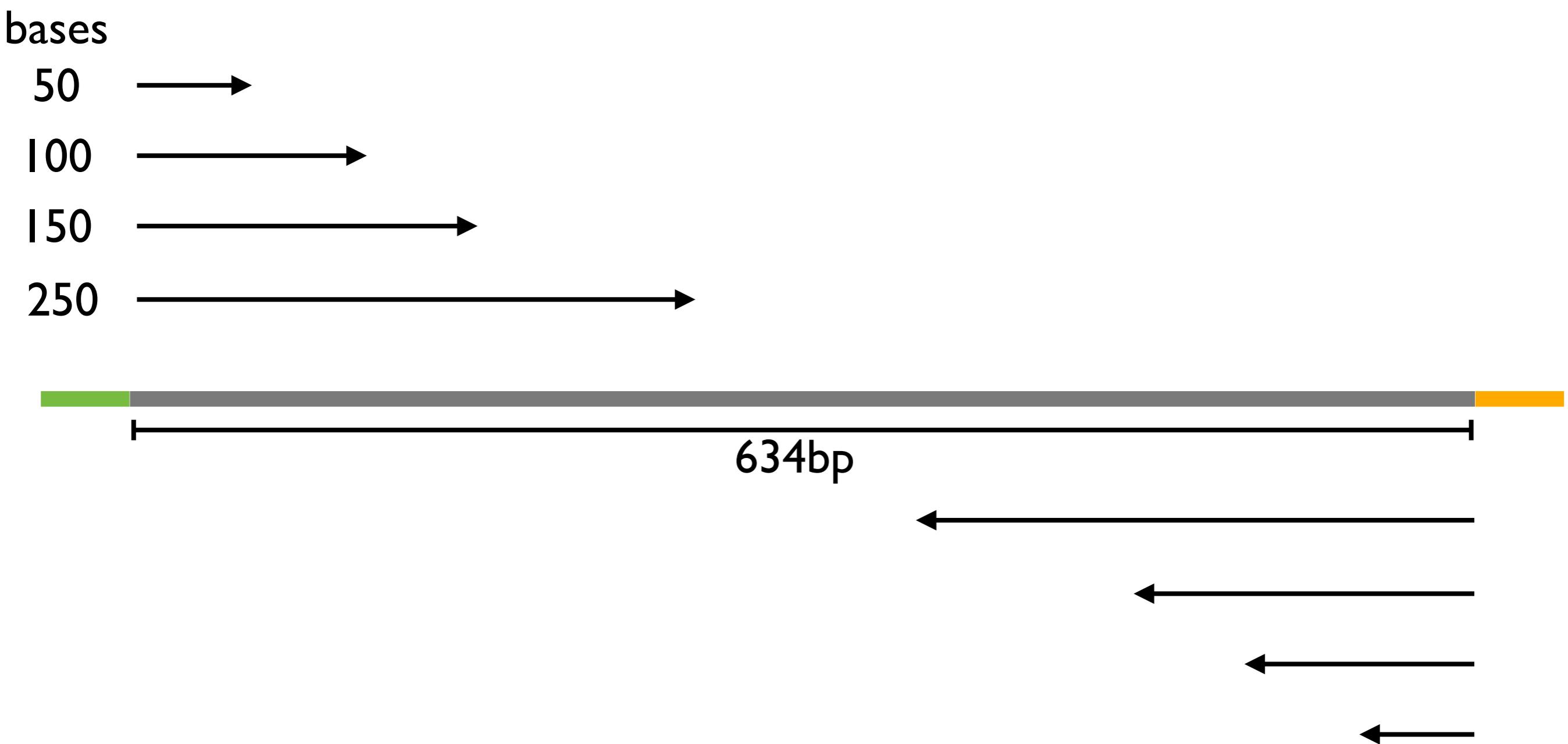
150 →



# Read Length



# Read Length



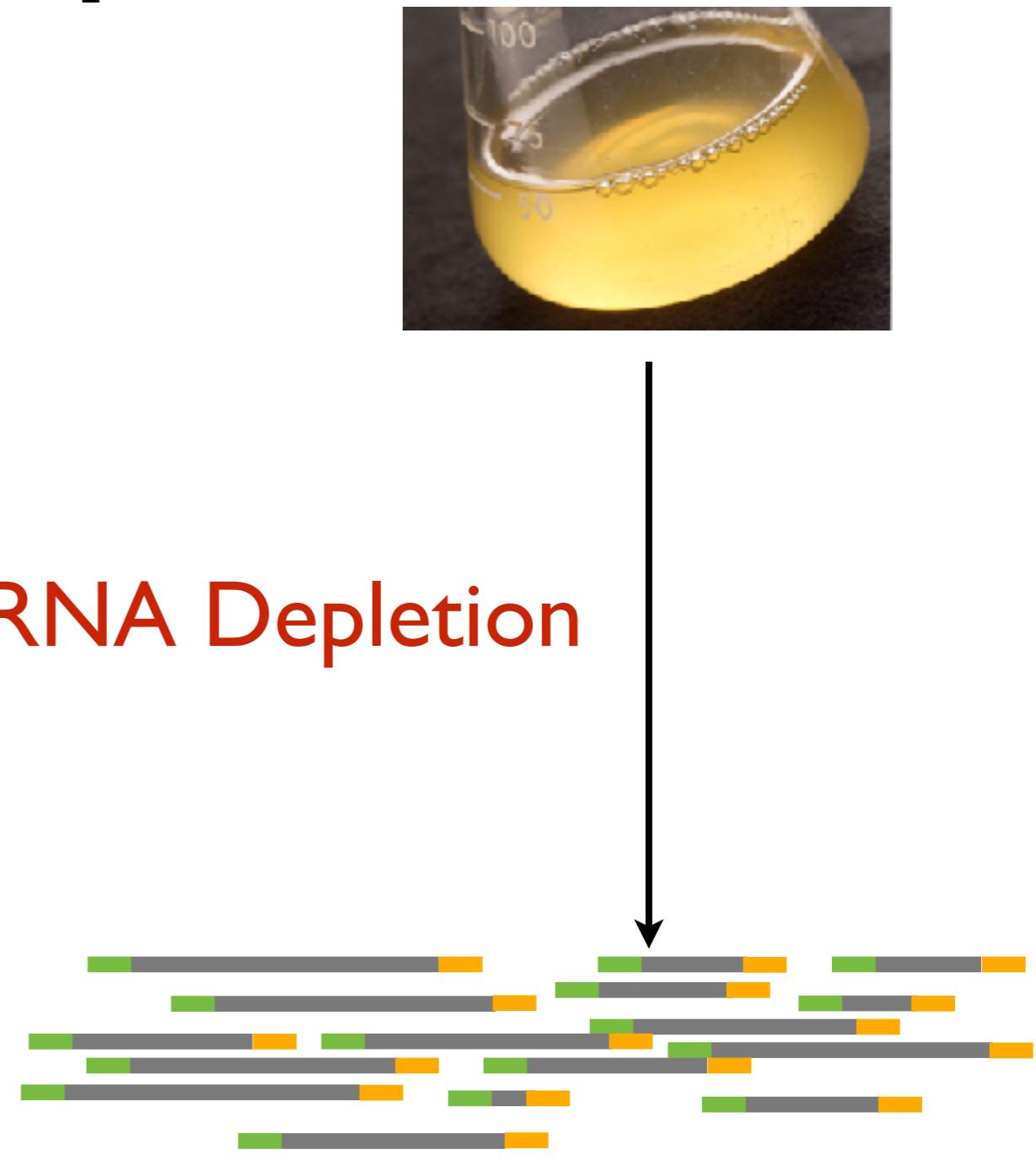
# MiSeq, NextSeq, and More Seqs

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

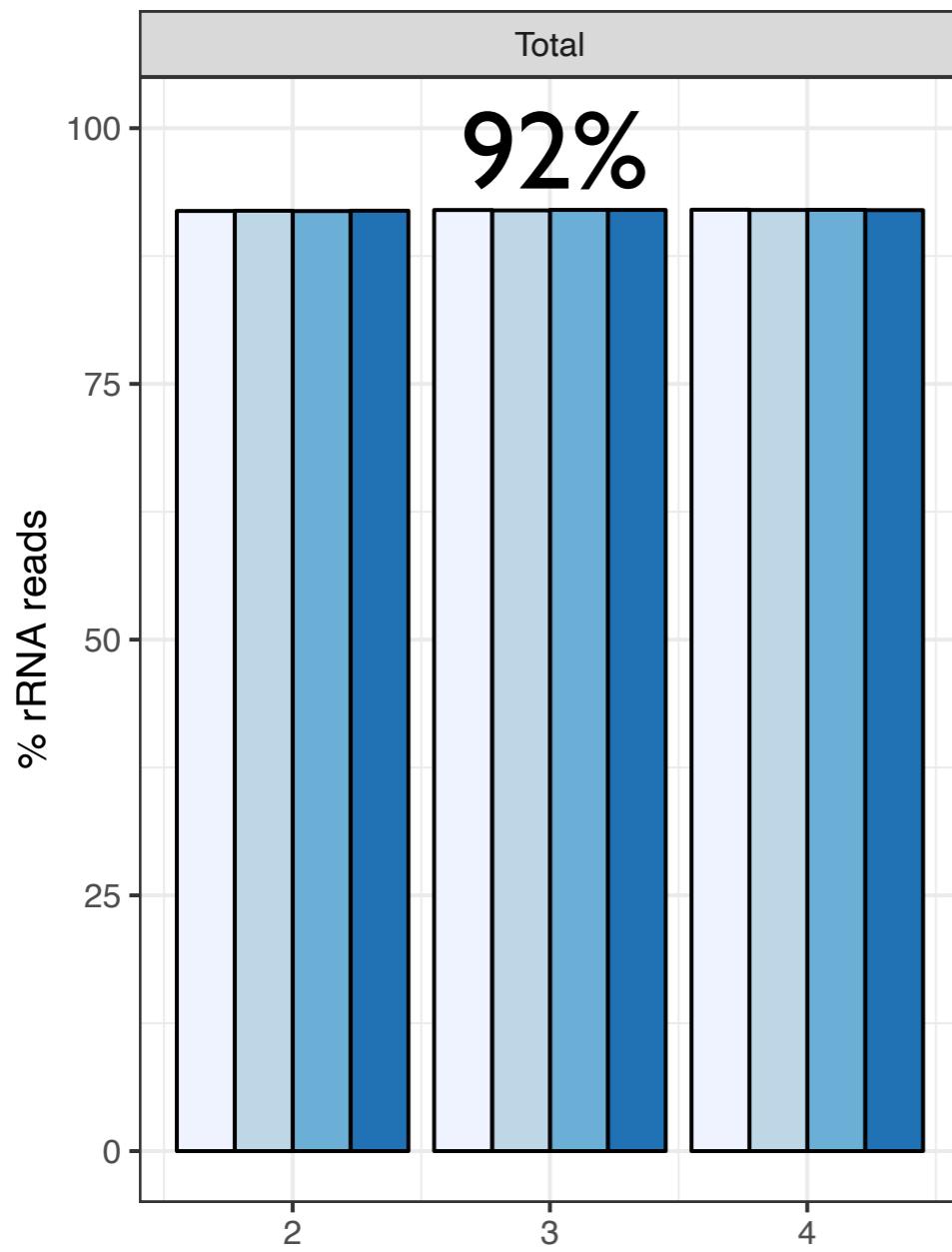
\* Duke Sequencing and Genomic Technologies Shared Resource, July 2018

# RNA-Seq: Major Components

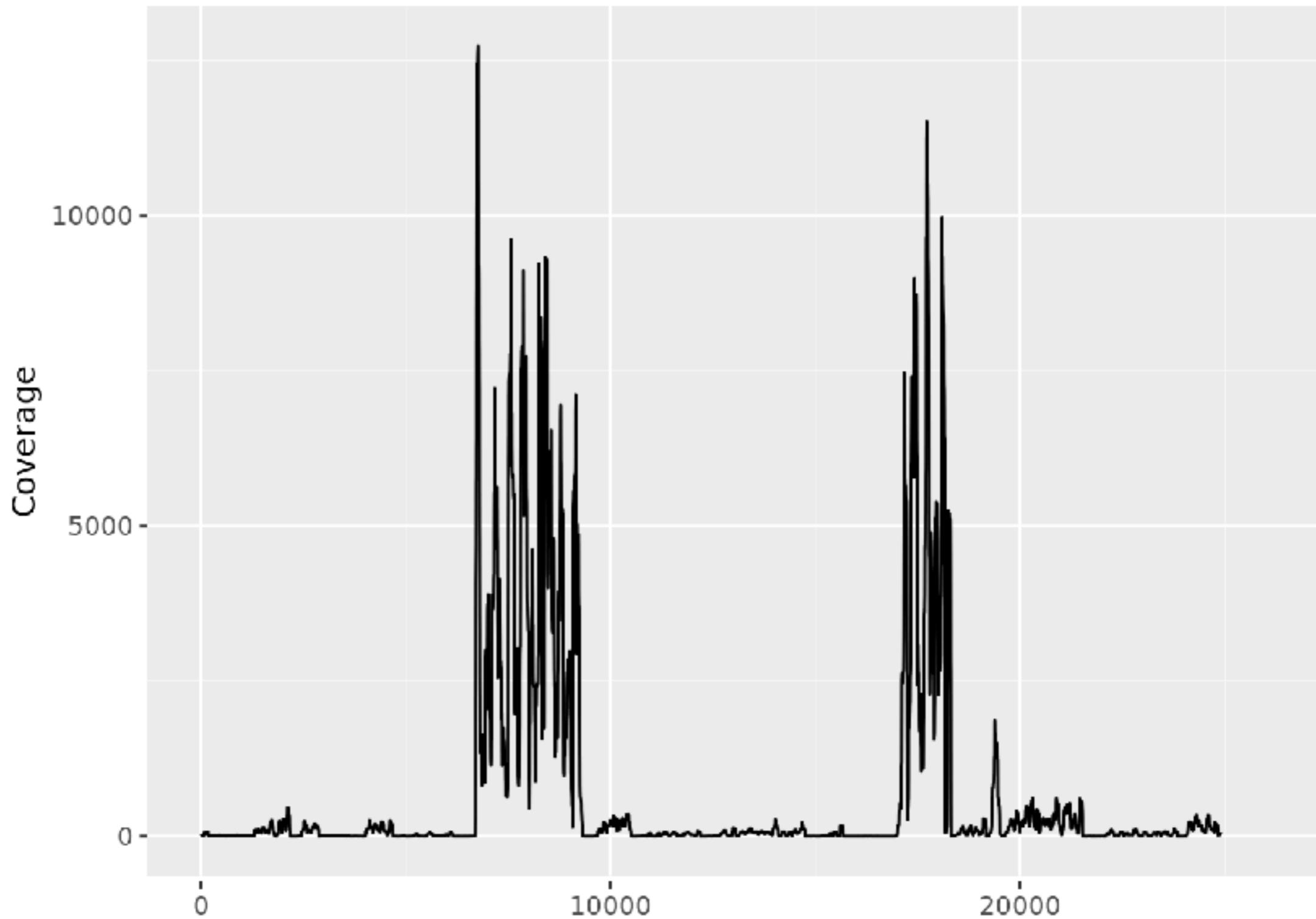
1. Sample Collection
2. RNA Extraction
3. mRNA Enrichment/rRNA Depletion
4. Library Preparation



# rRNA Depletion: Why?



# rRNA Depletion: Why?



# rRNA Depletion: How?

- Selection for desired RNA
  - poly(A) mRNA enrichment
  - Selective polyadenylation of mRNAs
  - Antibody capture of RNAs that interact with a specific protein
  - Non-random priming
- Selection against non-desired RNA
  - DNA targeted RNaseH degradation of rRNA
  - Ribosomal RNA capture
  - Duplex-specific nuclease (DSN) normalization
  - Degradation of processed RNA

# rRNA Depletion: How?

- Selection for desired RNA
  - poly(A) mRNA enrichment
    - Selective polyadenylation of mRNAs
    - Antibody capture of RNAs that interact with a specific protein
    - Non-random priming
  - Selection against non-desired RNA
    - DNA targeted RNaseH degradation of rRNA
    - Ribosomal RNA capture
    - Duplex-specific nuclease (DSN) normalization
    - Degradation of processed RNA

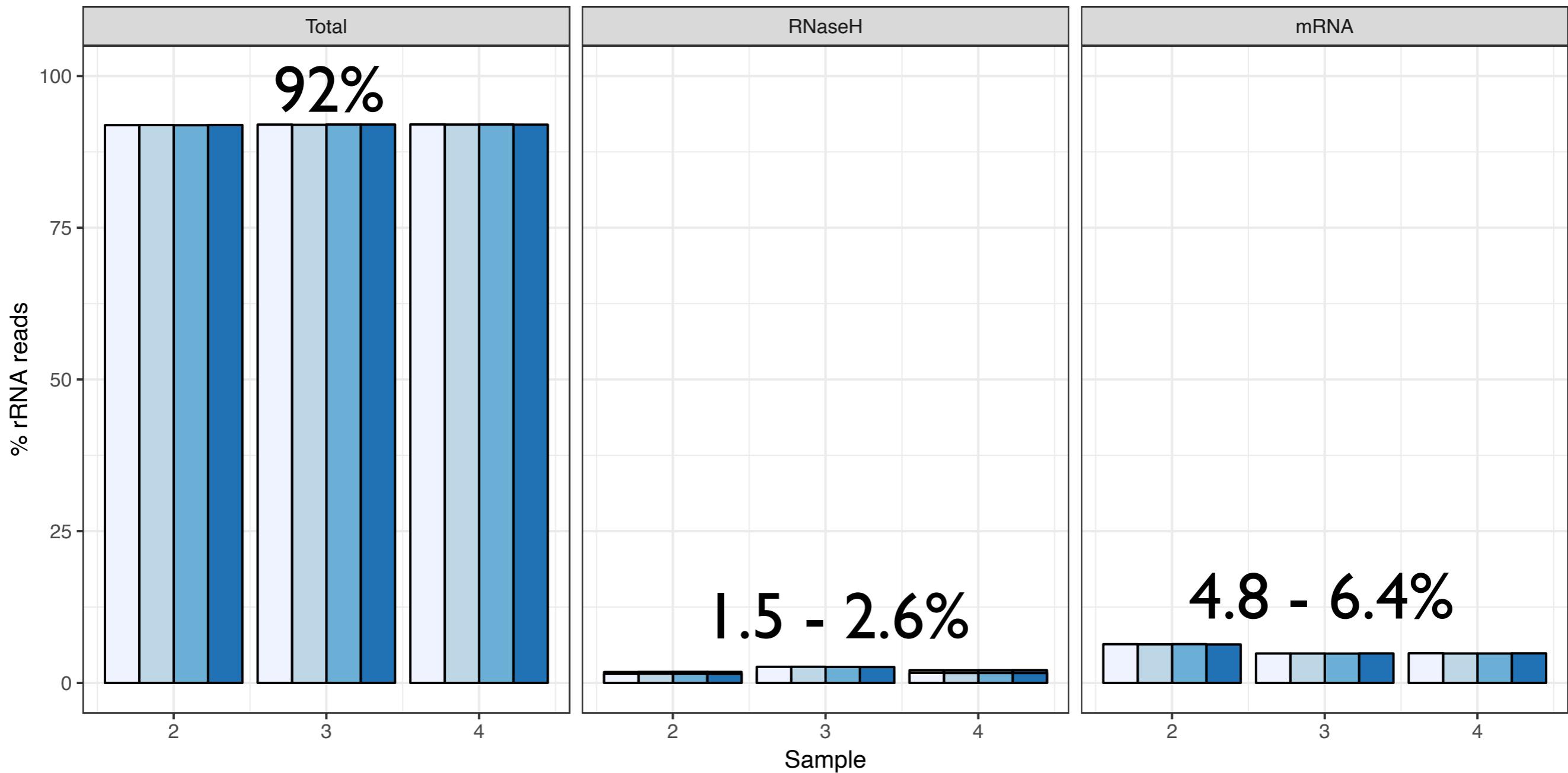
# rRNA Depletion: How?

- Selection for desired RNA
  - poly(A) mRNA enrichment
  - Selective polyadenylation of mRNAs
  - Antibody capture of RNAs that interact with a specific protein
  - Non-random priming
- Selection against non-desired RNA
  - DNA targeted RNaseH degradation of rRNA
  - Ribosomal RNA capture
  - Duplex-specific nuclease (DSN) normalization
  - Degradation of processed RNA

# rRNA Depletion: How?

- Selection for desired RNA
  - poly(A) mRNA enrichment
    - Selective polyadenylation of mRNAs
    - Antibody capture of RNAs that interact with a specific protein
    - Non-random priming
  - Selection against non-desired RNA
    - DNA targeted RNaseH degradation of rRNA
    - Ribosomal RNA capture
    - Duplex-specific nuclease (DSN) normalization
    - Degradation of processed RNA

# rRNA Depletion: How?

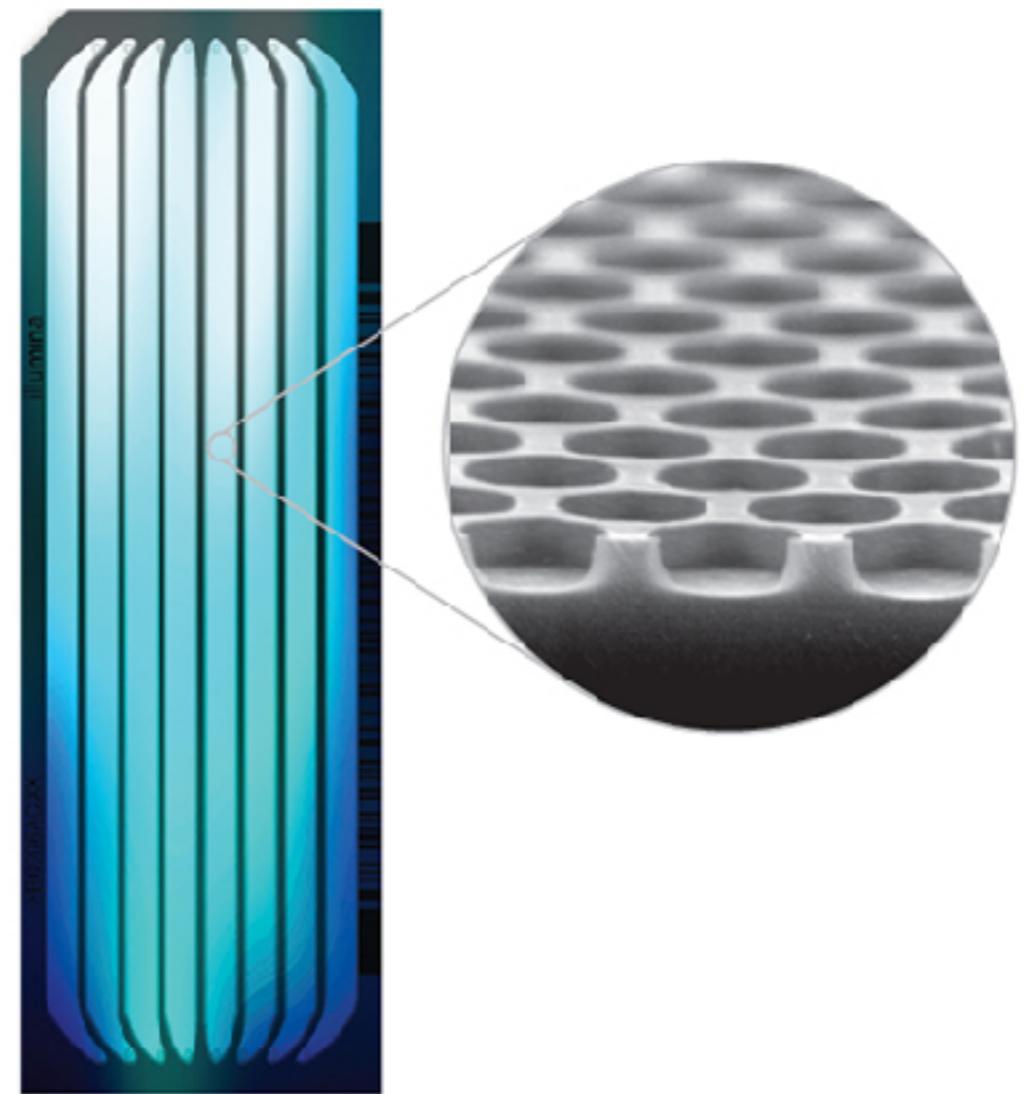


# Illumina Video

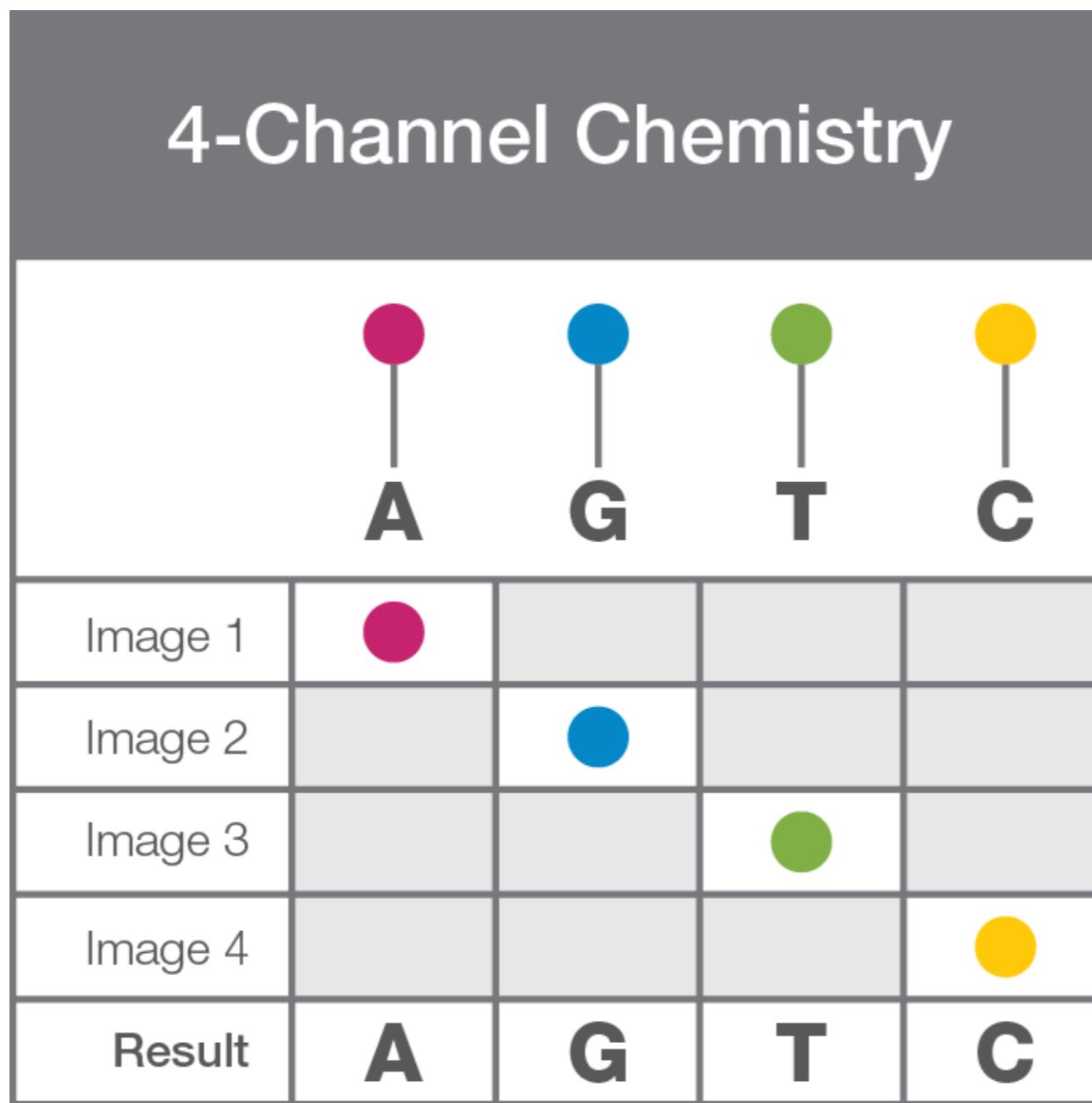
<https://www.youtube.com/watch?v=HMyCqWhwB8E>

# Patterned Flow Cells

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



# 4-Channel Chemistry



# Why Long Reads?

- Structural Variation
  - Large Insertions or Deletions
  - Duplications
  - Translocations
- De Novo Genome Assembly
- Phasing

# 2-Channel Chemistry

## 4-Channel Chemistry

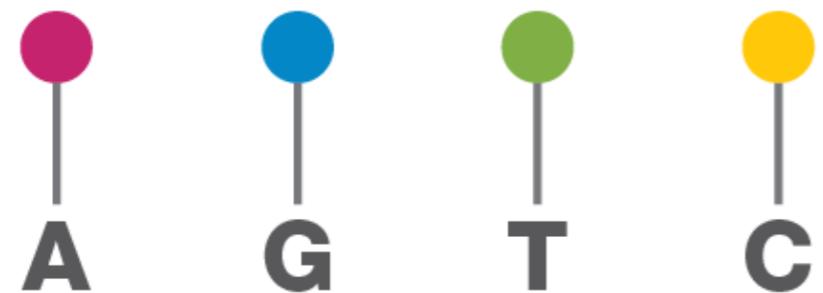


Image 1



Image 2



Image 3



Image 4



Result

**A**

**G**

**T**

**C**

## 2-Channel Chemistry



Image 1



Image 2



Result

**A**

**G**

**T**

**C**

# Comparing Technologies

<b>Method</b>	<b>Read length</b>	<b>Accu racy</b>	<b>Reads per run</b>	<b>Max Output</b>	<b>Cost (\$/Mb)</b>	<b>Pros</b>	<b>Cons</b>
Sanger	400-900 bp	99.9%	1	900 bp	\$2400	Longer reads.	Expensive. Low Output
Illumina	600 bp (300bp PE)	99.9%	$20 \times 10^9$	6000 Gb	\$0.01	High yield per base cost	Equipment expense. Short reads
PacBio	>10kb ave. >40kb max	99%	$5 \times 10^5$	10 Gb	\$0.08	Very long reads	Homopolymer errors. Moderate Output. Equipment expense.
Nanopore	>100 kb N50 >1Mb Max	92%	$1 \times 10^6$	5 Gb	\$0.10	Very long reads Portable Cheap Equipment	Homopolymer errors. Moderate Output.

# Single Molecule Technologies

# DNA Sequencing Technologies (Abridged)

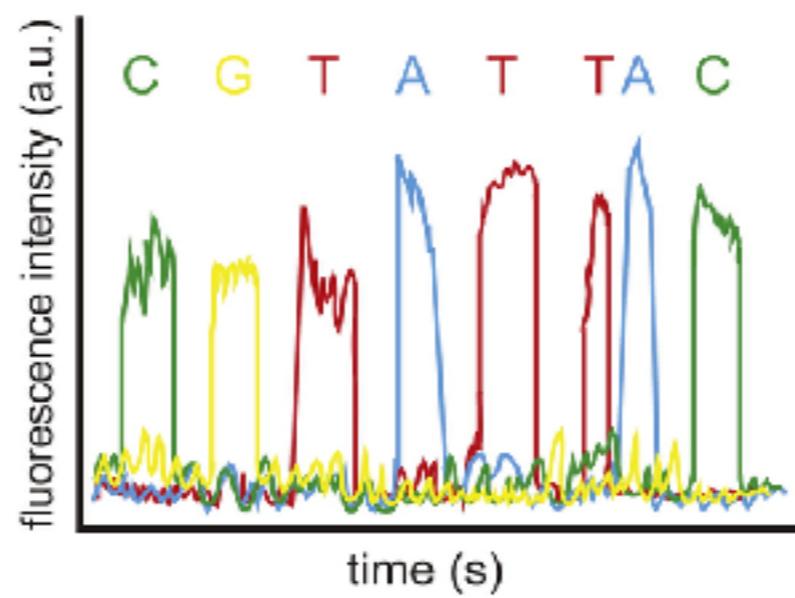
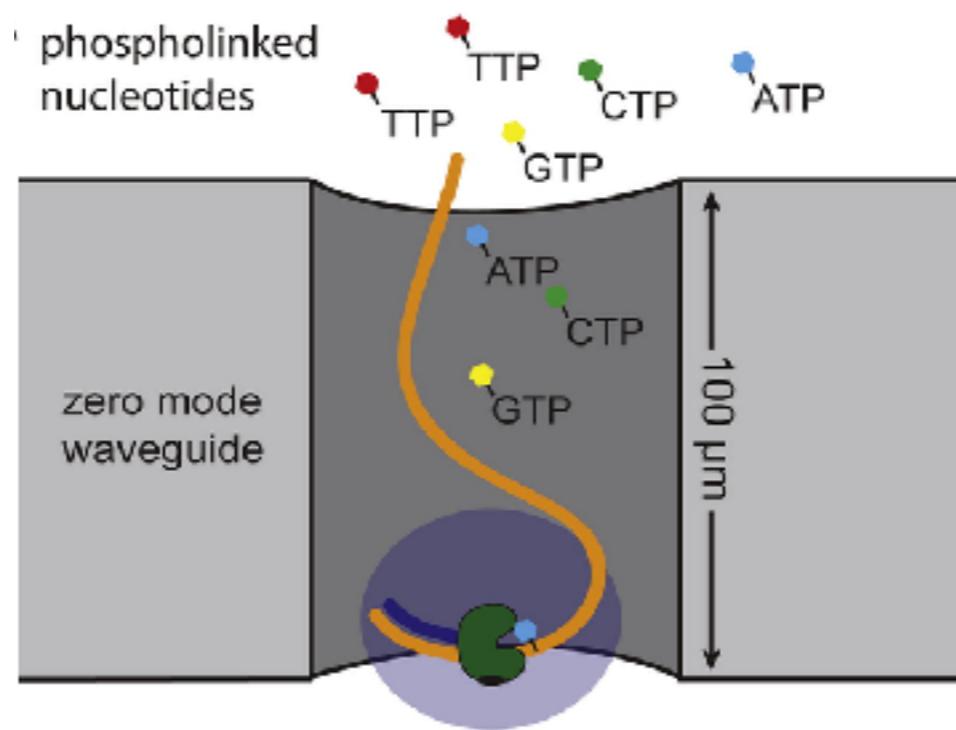
1st Generation	2nd Generation	3rd Generation
Chemical (Maxim-Gilbert)	Pyrosequencing (454)	Single molecule real time (PacBio)
Chain Termination (Sanger)	Chain Termination (Illumina)	Nanopore sequencing (Oxford Nanopore)
Pyrosequencing	Sequencing by ligation (SOLiD sequencing)	
	Ion semiconductor (Ion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

# Sequencing by Synthesis

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)	

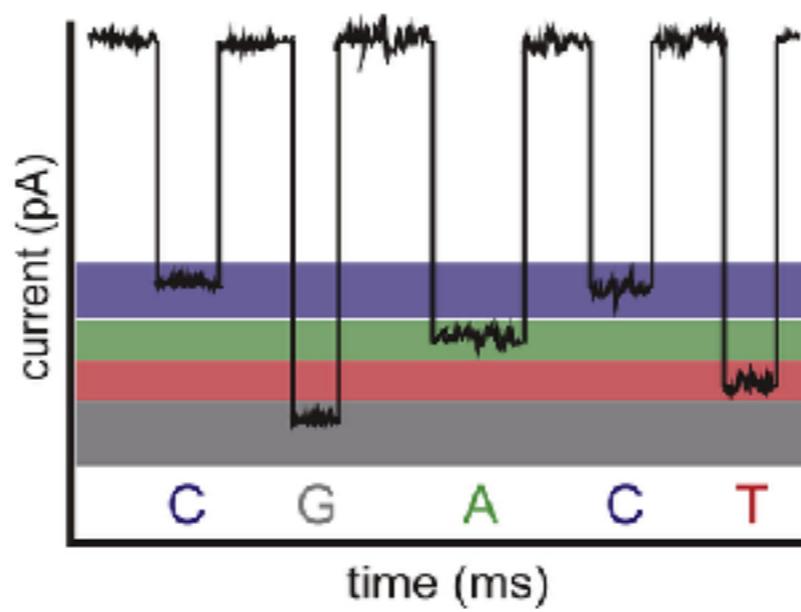
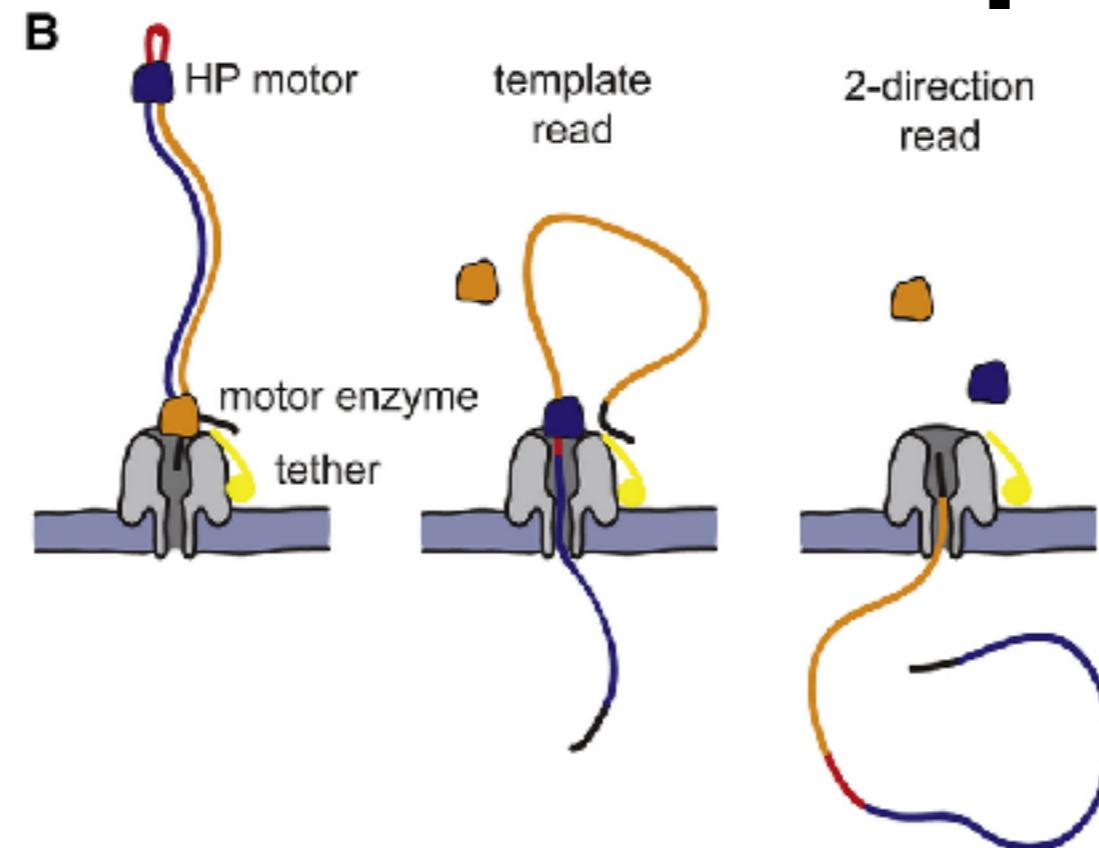
1st Generation	2nd Generation	3rd Generation
Chemical (Maxim-Gilbert)	Pyrosequencing (454)	Single molecule real time (PacBio)
Chain Termination (Sanger)	Chain Termination (Illumina)	Nanopore sequencing (Oxford Nanopore)
Pyrosequencing	Sequencing by ligation (SOLID sequencing)	
	Ion semiconductor (Ion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

# Pacific Biosciences



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)	

# Oxford Nanopore



# Sequencers



<https://www2.nanoporetech.com/images/product-page/MinION-Banner.jpg>  
<http://www.gatc-biotech.com/en/gatc/sequencing-technologies/pacbio-rs-ii.html>  
<http://www.dnavision.com/illumina.php>