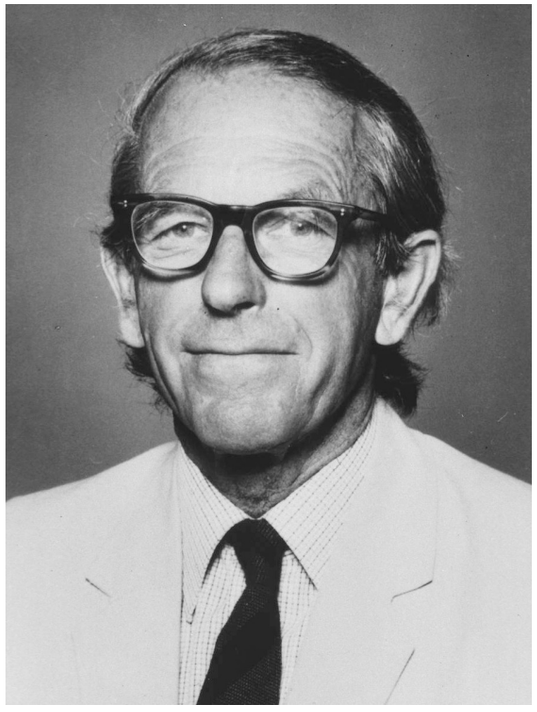
CSE 566 Spring 2023

Sequencing Technologies

Instructor: Mingfu Shao

Sequencing Technologies

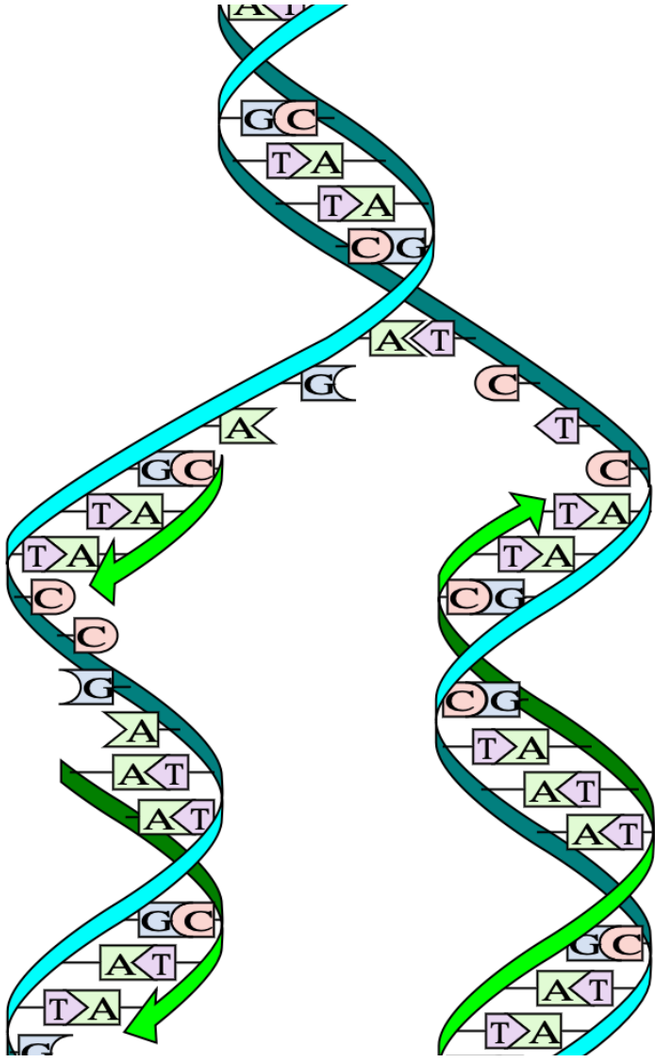
- 1st generation sequencing
  - Sanger sequencing (1977)
- 2nd/next generation sequencing (NGS)
  - Illumina sequencing
- 3rd generation sequencing
  - Pacific Biosciences (PacBio)
  - Oxford Nanopore Technologies (ONT)

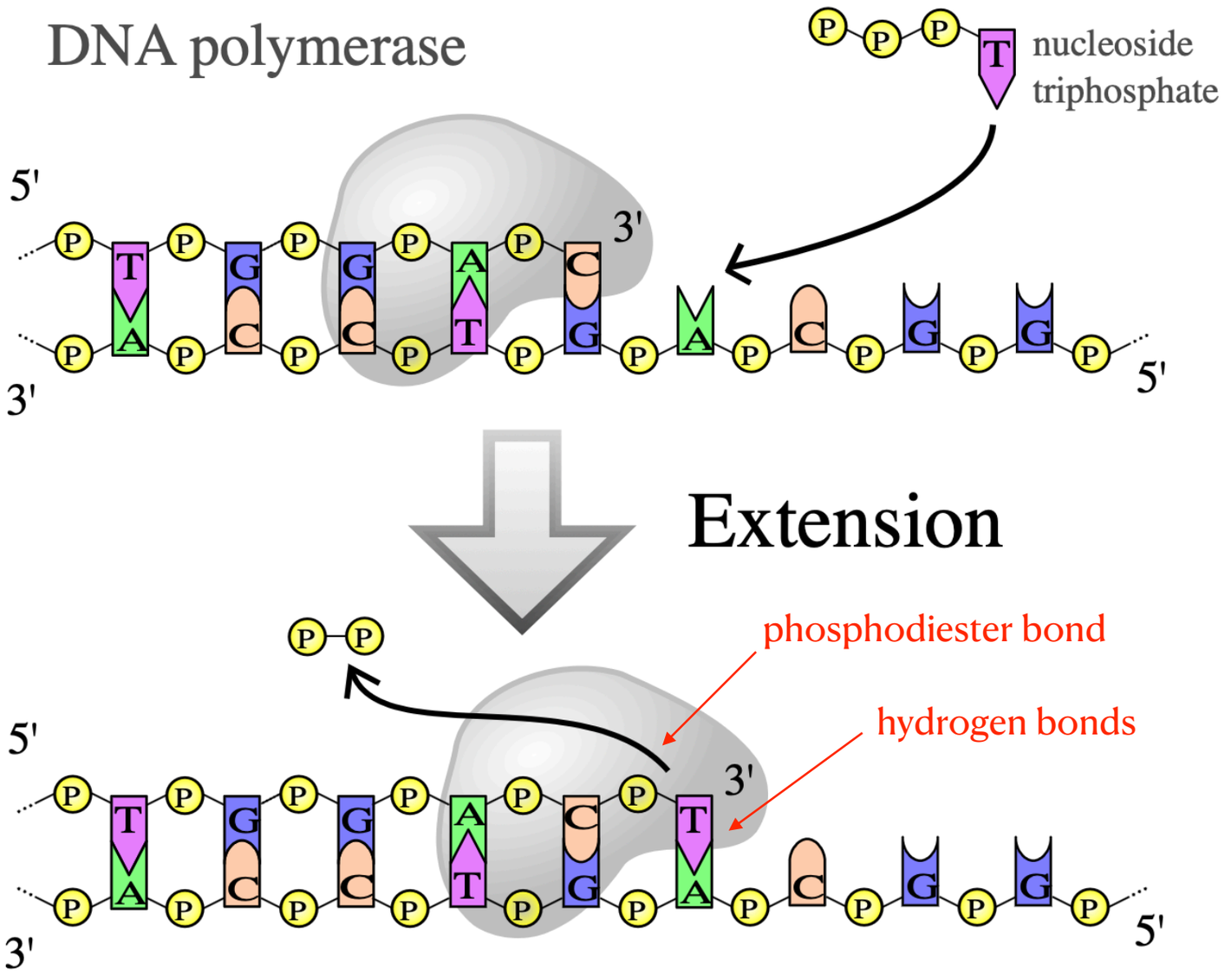


Frederick Sanger

1918—2013

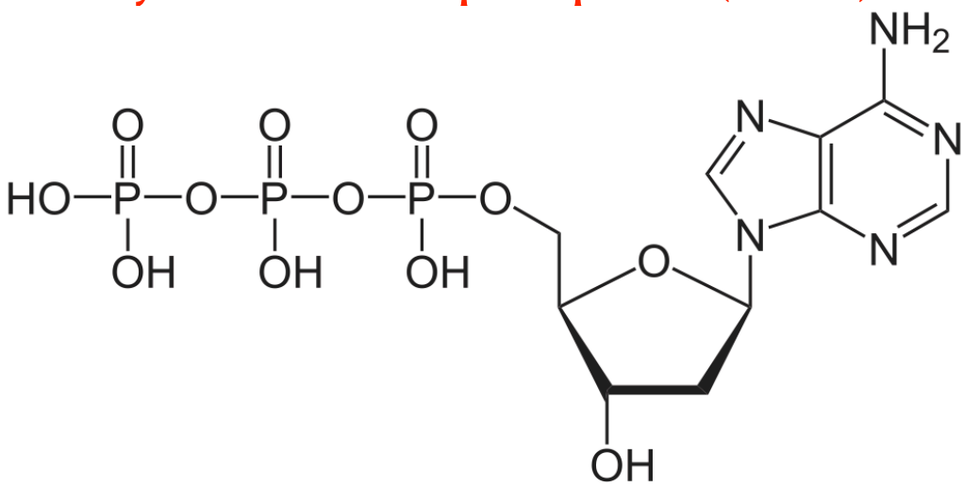
DNA Replication



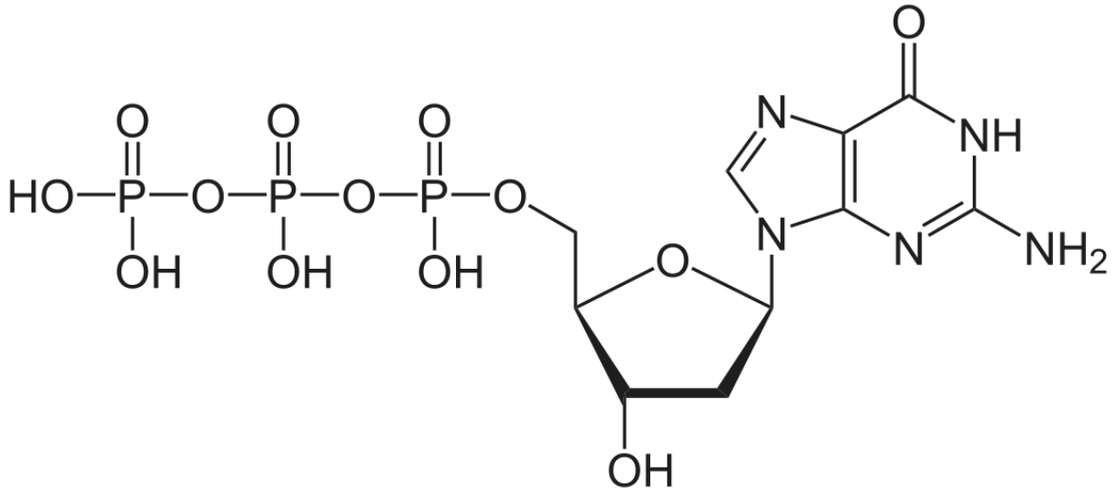


dNTPs

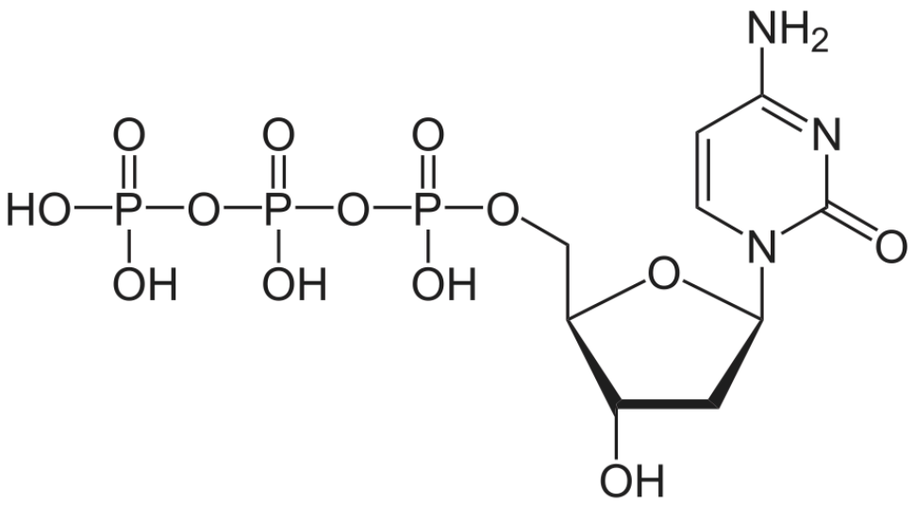
deoxyadenosine triphosphate (dATP)



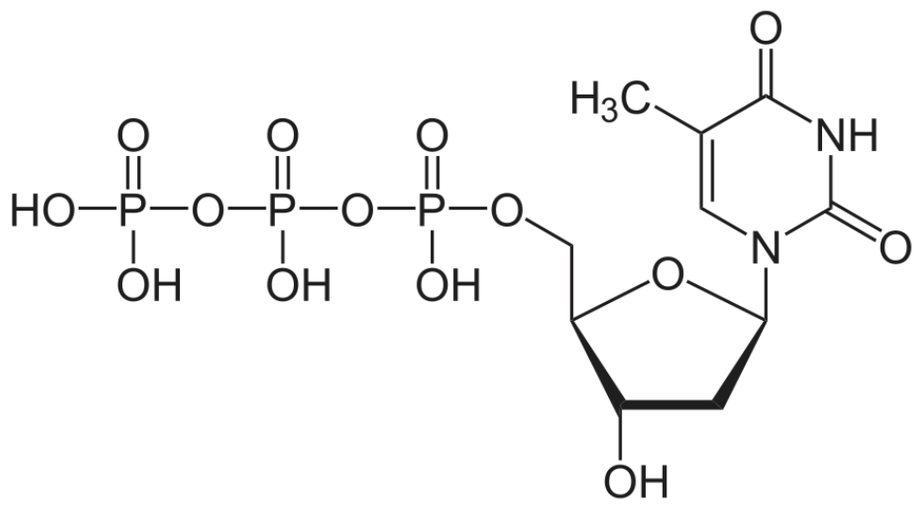
deoxyguanosine triphosphate (dGTP)



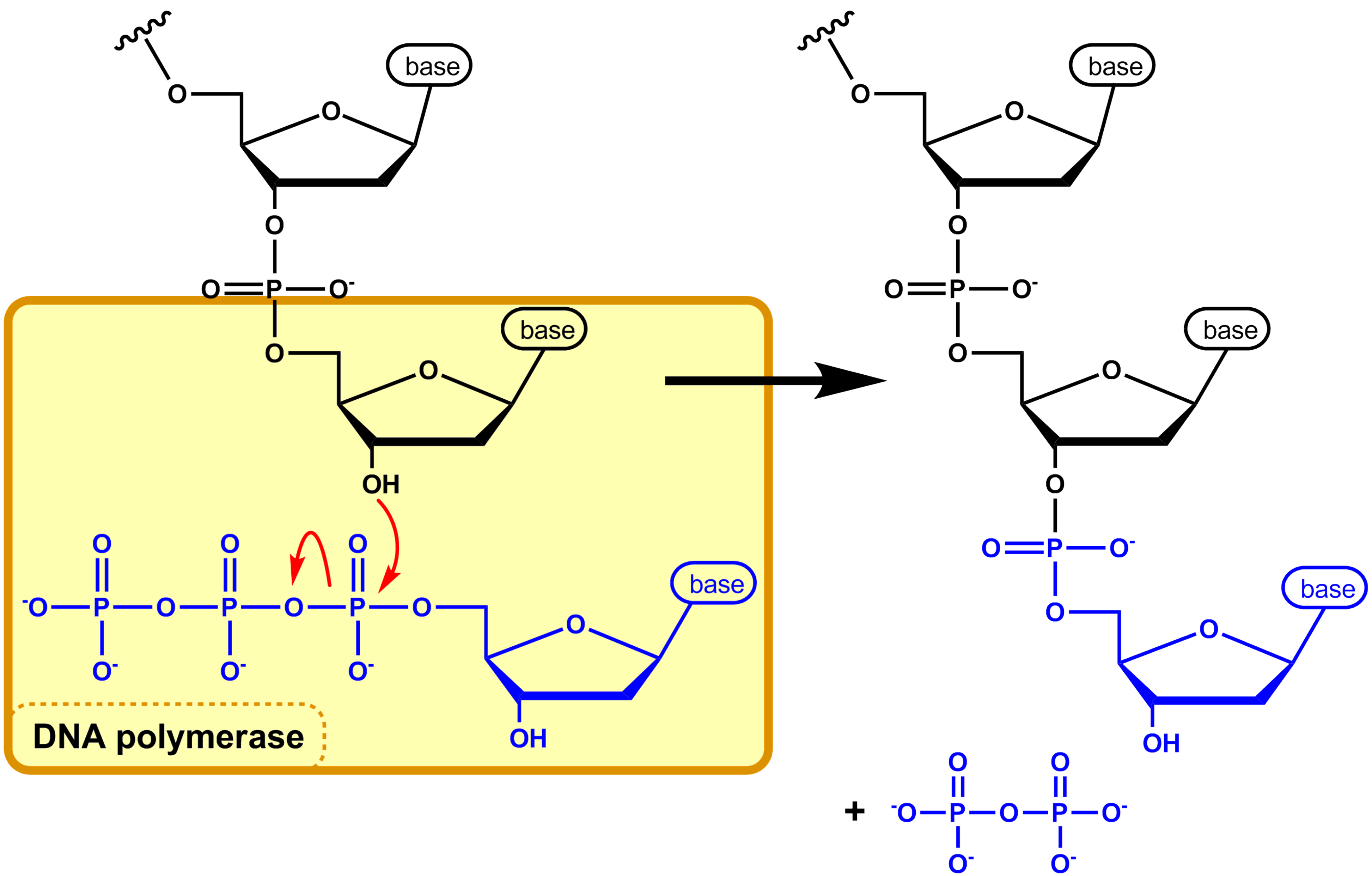
deoxycytidine triphosphate (dCTP)



deoxythymidine triphosphate (dTTP)

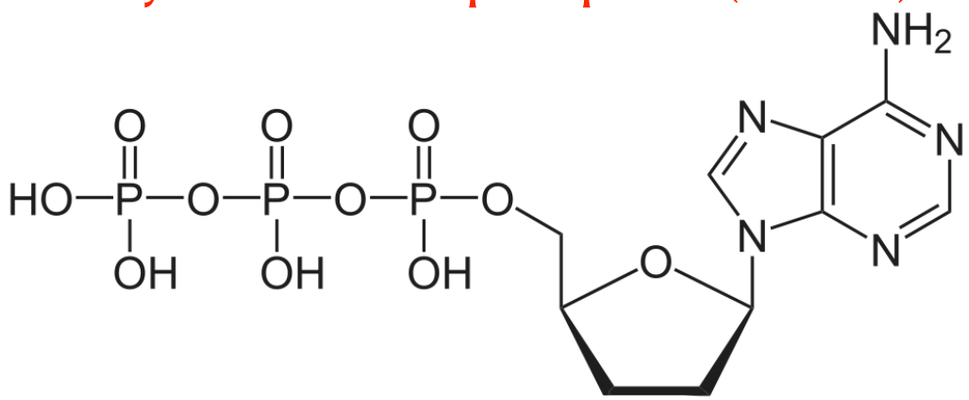


Forming Phosphodiester Bond

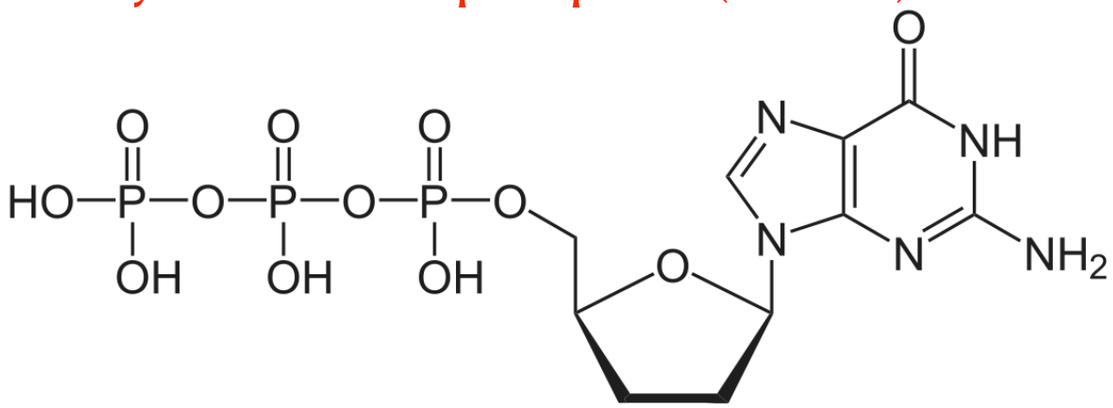


ddNTPs

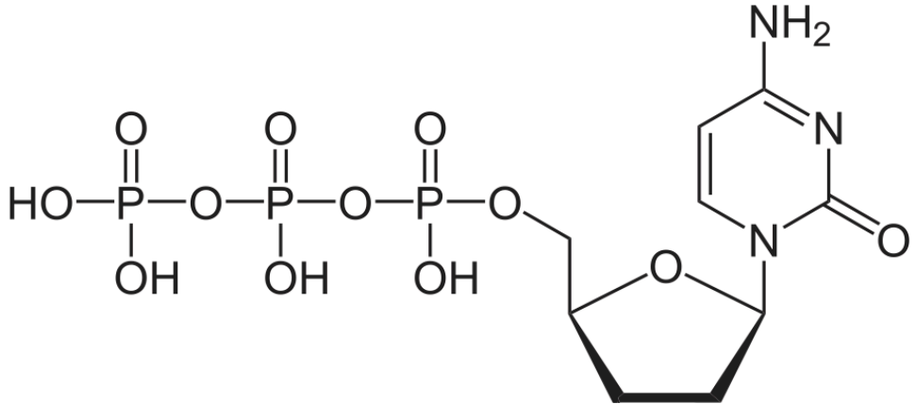
dideoxyadenosine triphosphate (ddATP)



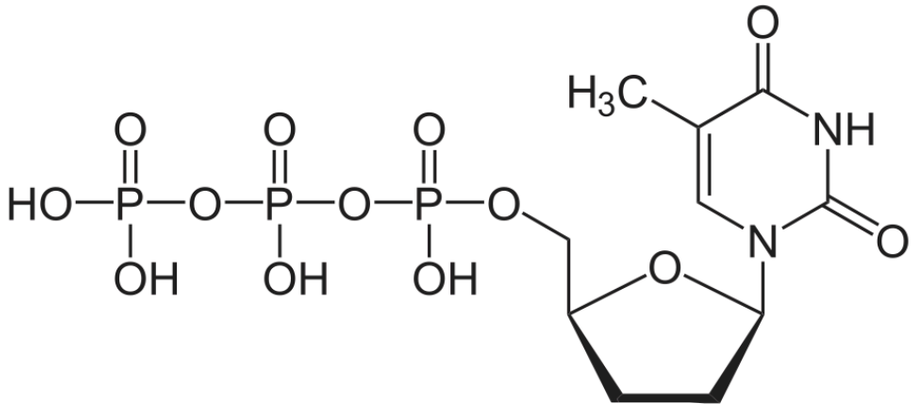
dideoxyadenosine triphosphate (ddGTP)



dideoxyadenosine triphosphate (ddCTP)



dideoxyadenosine triphosphate (ddTTP)



- ddNTPs can be added to the chain by the DNA polymerase, but will also terminate the synthesis.

Sanger Sequencing (version 1.0)

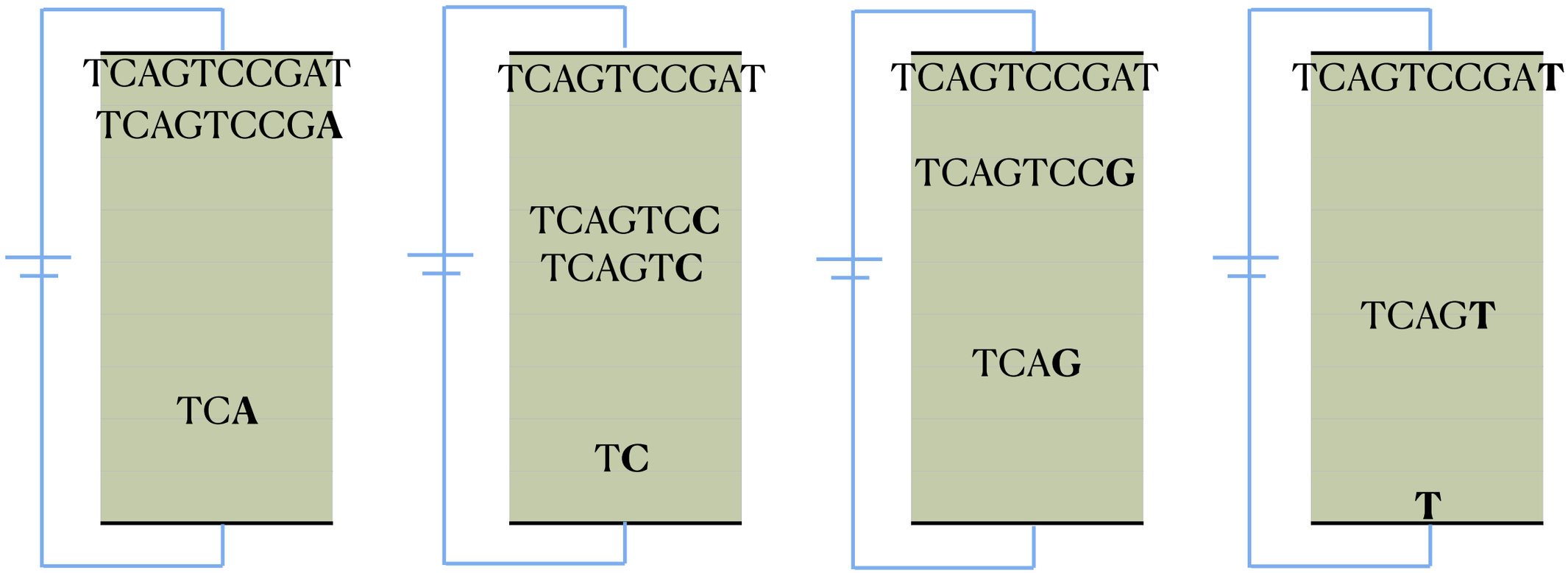
Tube A	Tube C	Tube G	Tube T
TCAGTCCGAT (targeted DNAs)	TCAGTCCGAT (targeted DNAs)	TCAGTCCGAT (targeted DNAs)	TCAGTCCGAT (targeted DNAs)
DNA polymerase	DNA polymerase	DNA polymerase	DNA polymerase
dATP dCTP dGTP dTTP	dATP dCTP dGTP dTTP	dATP dCTP dGTP dTTP	dATP dCTP dGTP dTTP
ddATP	**ddCTP**	**ddGTP**	**ddTTP**

Sanger Sequencing (version 1.0)

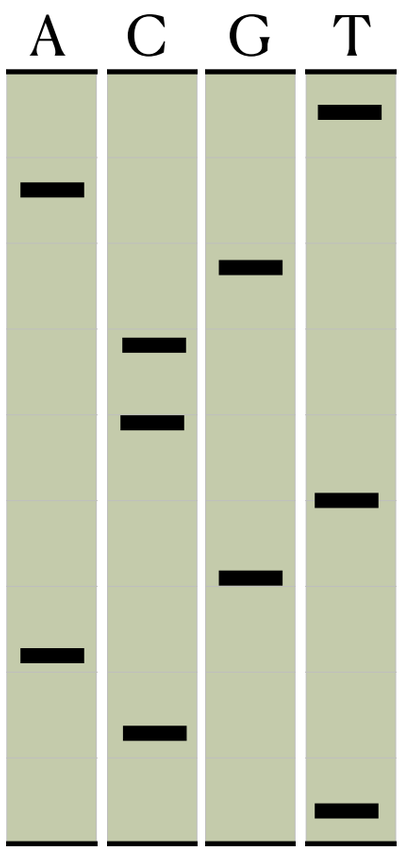
Tube A	Tube C	Tube G	Tube T
TCAGTCCGAT (targeted DNAs)	TCAGTCCGAT (targeted DNAs)	TCAGTCCGAT (targeted DNAs)	TCAGTCCGAT (targeted DNAs)
DNA polymerase	DNA polymerase	DNA polymerase	DNA polymerase
dATP dCTP dGTP dTTP **ddATP**	dATP dCTP dGTP dTTP **ddCTP**	dATP dCTP dGTP dTTP **ddGTP**	dATP dCTP dGTP dTTP **ddTTP**
TCAGTCCG**A** TCAGTCCG**A** TC**A** TC**A** TC**A** TCAGTCCGAT	TCAGTC**C** T**C** T**C** TCAGT**C** TCAGT**C** TCAGTCCGAT	TCAGTCC**G** TCAGTCC**G** TCA**G** TCA**G** TCA**G** TCAGTCCGAT	TCAGTCCGA**T** TCAG**T** **T** **T** TCAG**T** TCAG**T**

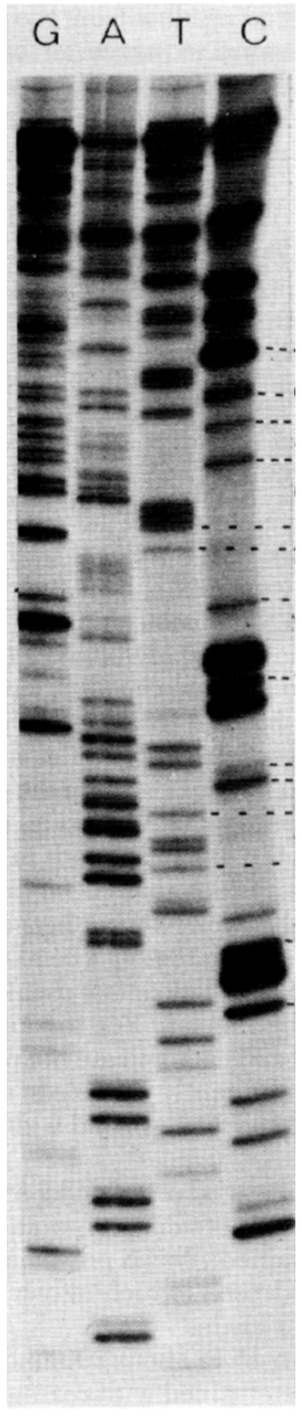
Gel Electrophoresis

Tube A	Tube C	Tube G	Tube T
TCAGTCCG**A**	TCAGTC**C**	TCAGTCC**G**	TCAGTCCGA**T**
TCAGTCCG**A**	T**C**	TCAGTCC**G**	TCAG**T**
TC**A**	T**C**	TCA**G**	**T**
TC**A**	TCAGT**C**	TCA**G**	**T**
TC**A**	TCAGT**C**	TCA**G**	TCAG**T**
TCAGTCCGAT	TCAGTCCGAT	TCAGTCCGAT	TCAG**T**



Gel Electrophoresis



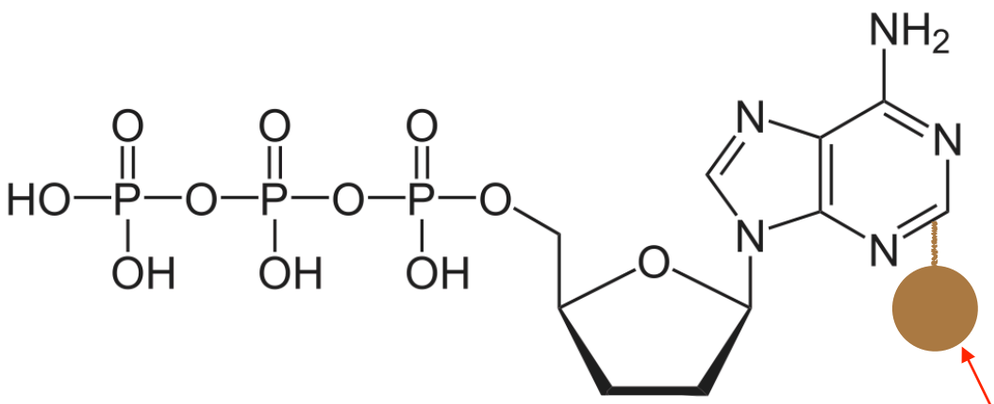


Sanger et al.,

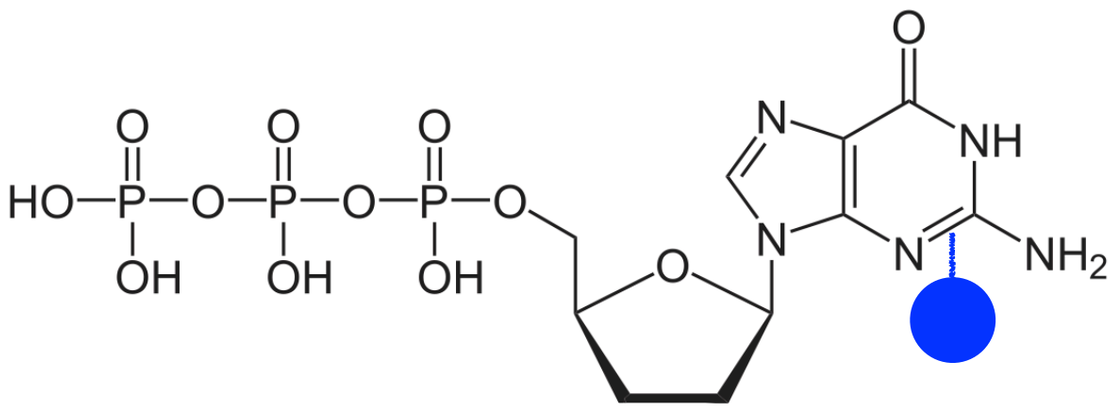
1977, *PNAS*

Fluorescent ddNTPs

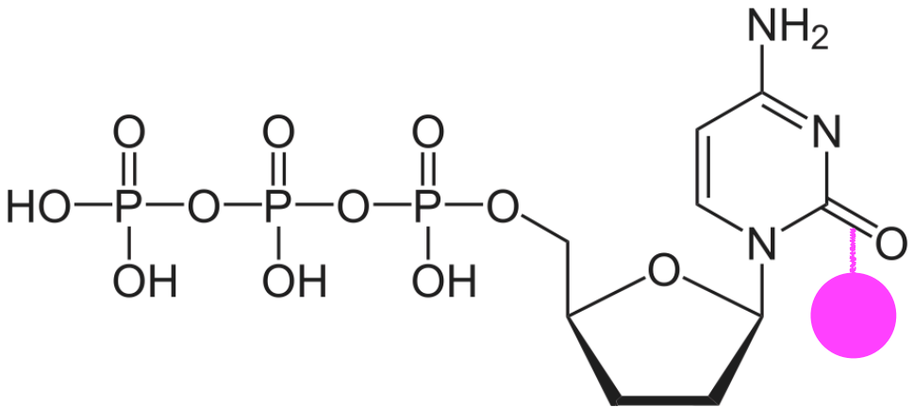
Fluorescent ddATP



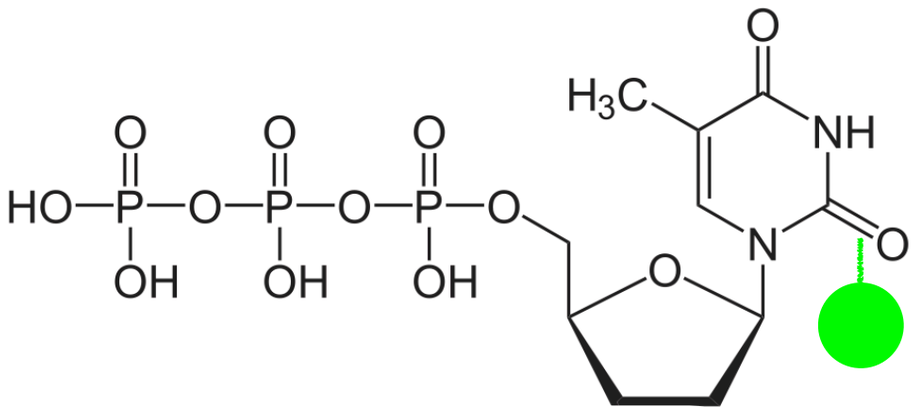
Fluorescent ddGTP



Fluorescent ddCTP

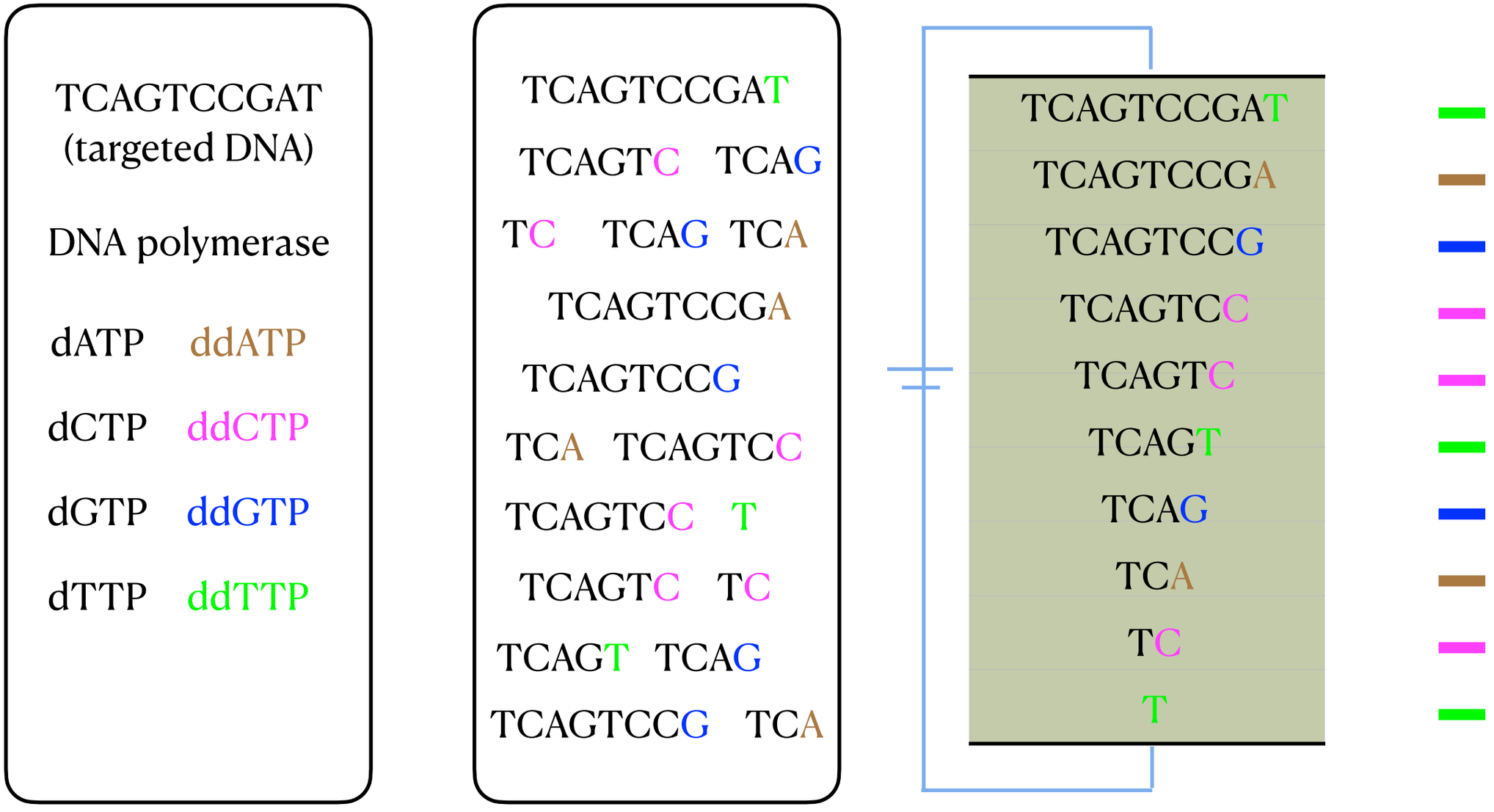


Fluorescent ddTTP



- Fluorescent ddNTPs can be “read-out” by an optical detector.

Sanger Sequencing (version 2.0)

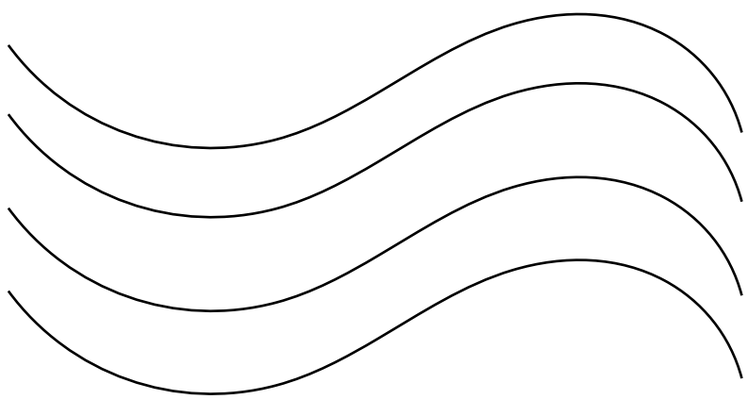


Sanger Sequencing

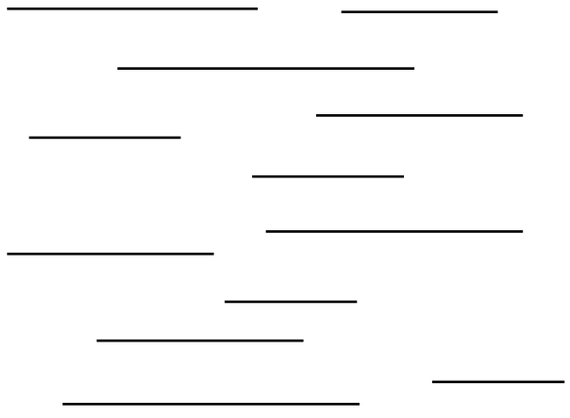
- Open a new era
  - First bacterial genome sequenced (1995)
  - First draft of human genome (2001)
- Build a technical foundation for 2nd/3rd-gen.
- Limitations
  - Low-quality sequences within the first 15-40 bps
  - Inability to distinguish after ~800 bps
  - Low throughput

2nd/Next Generation Sequencing

1. Short-gun sequencing, massively parallel
2. Paired-end, short reads
3. Highly accurate: error rate < 1%
4. Represented by Illumina products



Many copies of DNA

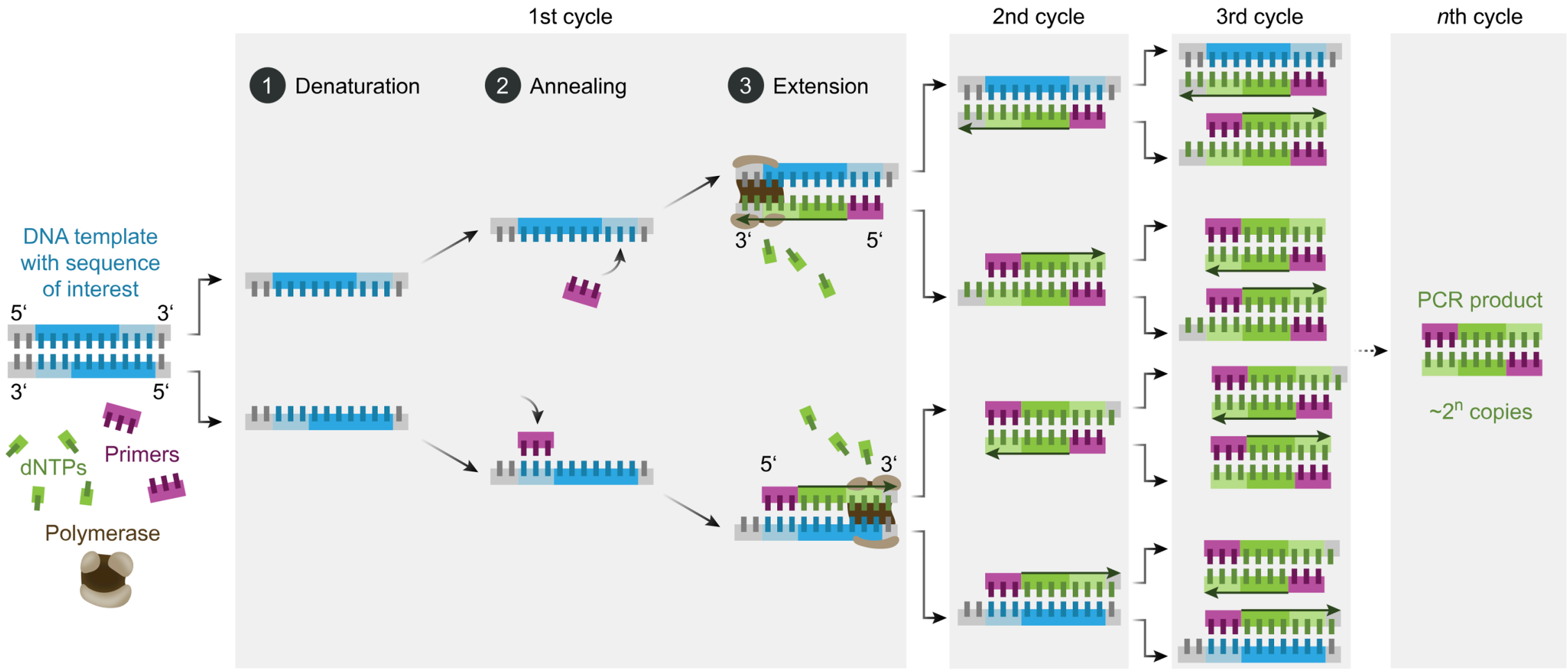


Many shorter fragments

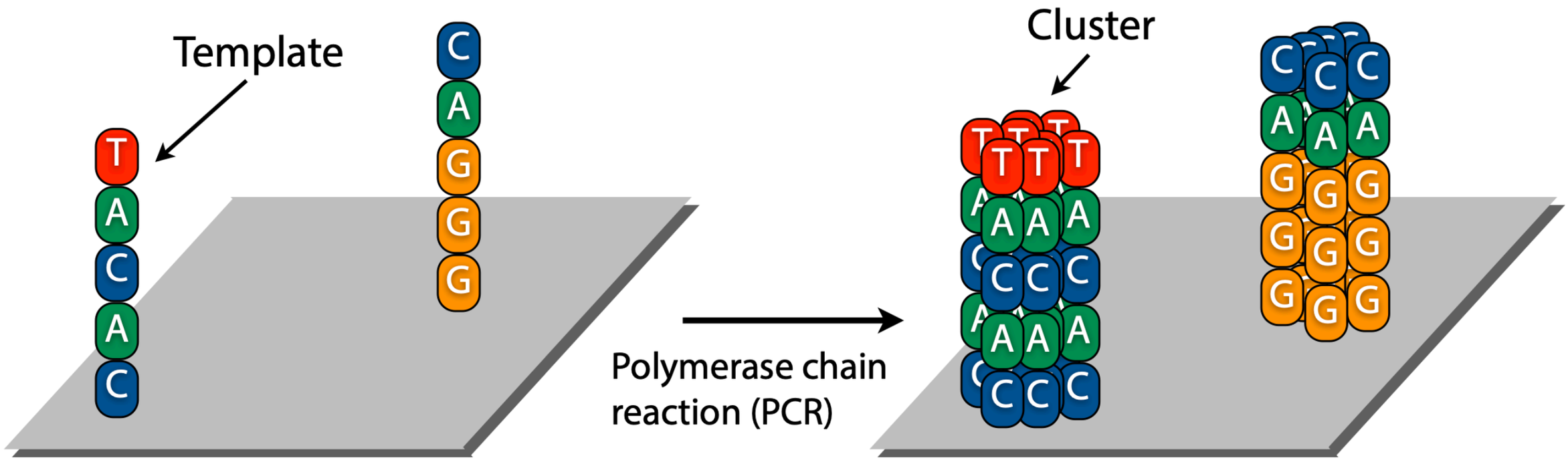
Paired-end reads

PCR (Polymerase Chain Reaction)

- For rapid amplification of DNAs



Bridge PCR

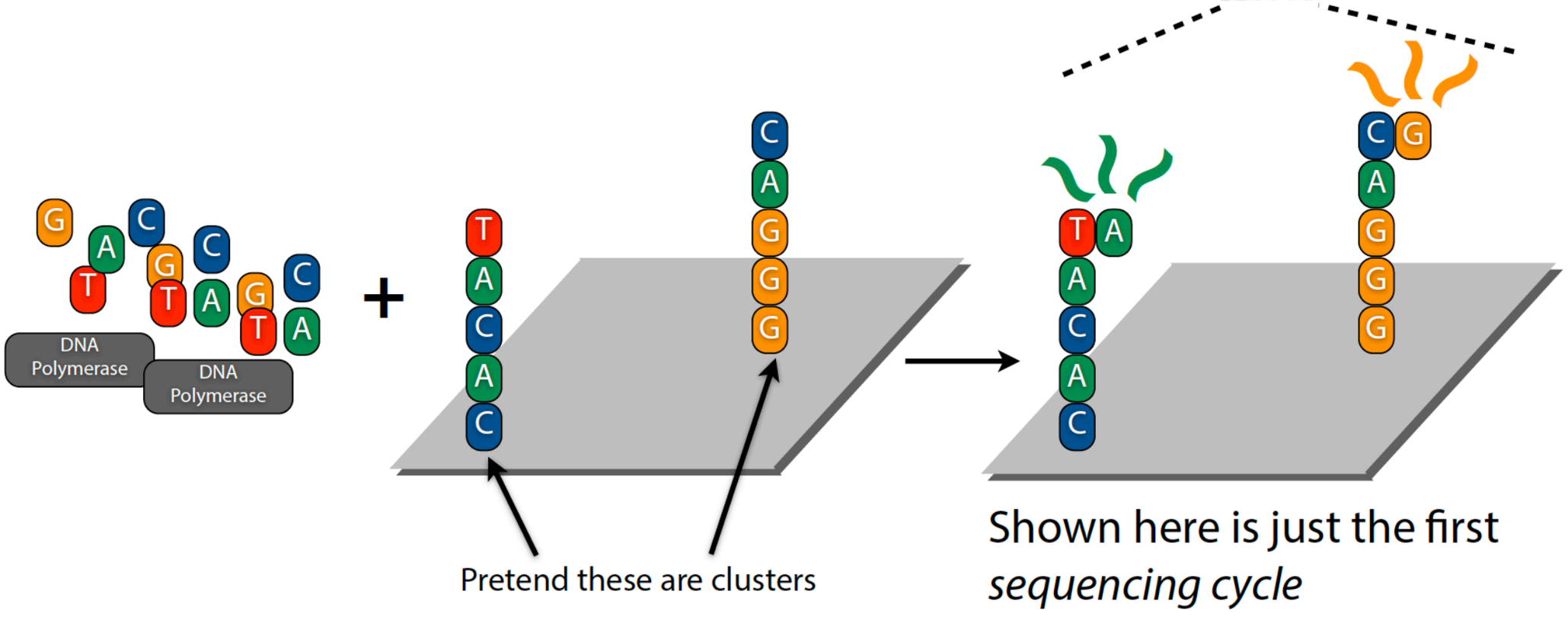


- Outcomes: tens of thousands of “clusters of fragments”, each of which share the same sequence.

Sequencing by Synthesis (SBS)

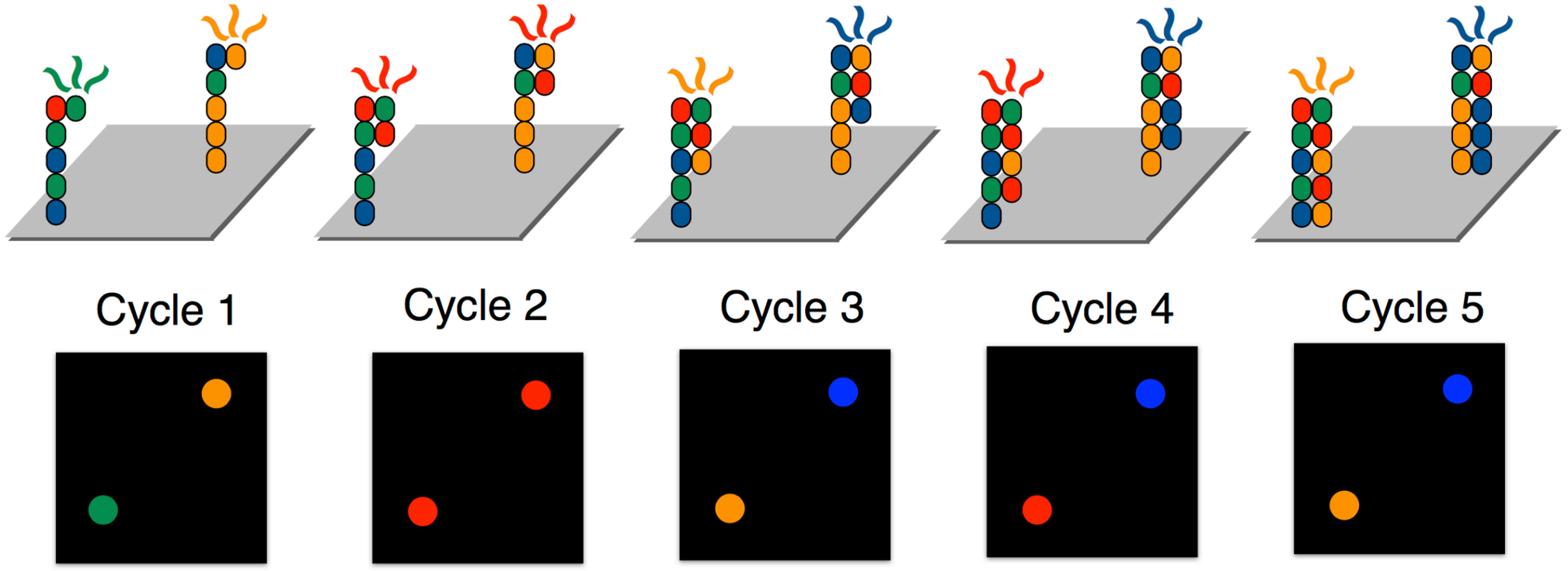
- In each cycle, DNA polymerase and fluorescent dNTPs are added, a single nucleotide will be synthesized in every template, and all fluorescent signals will be captured.

(snap)

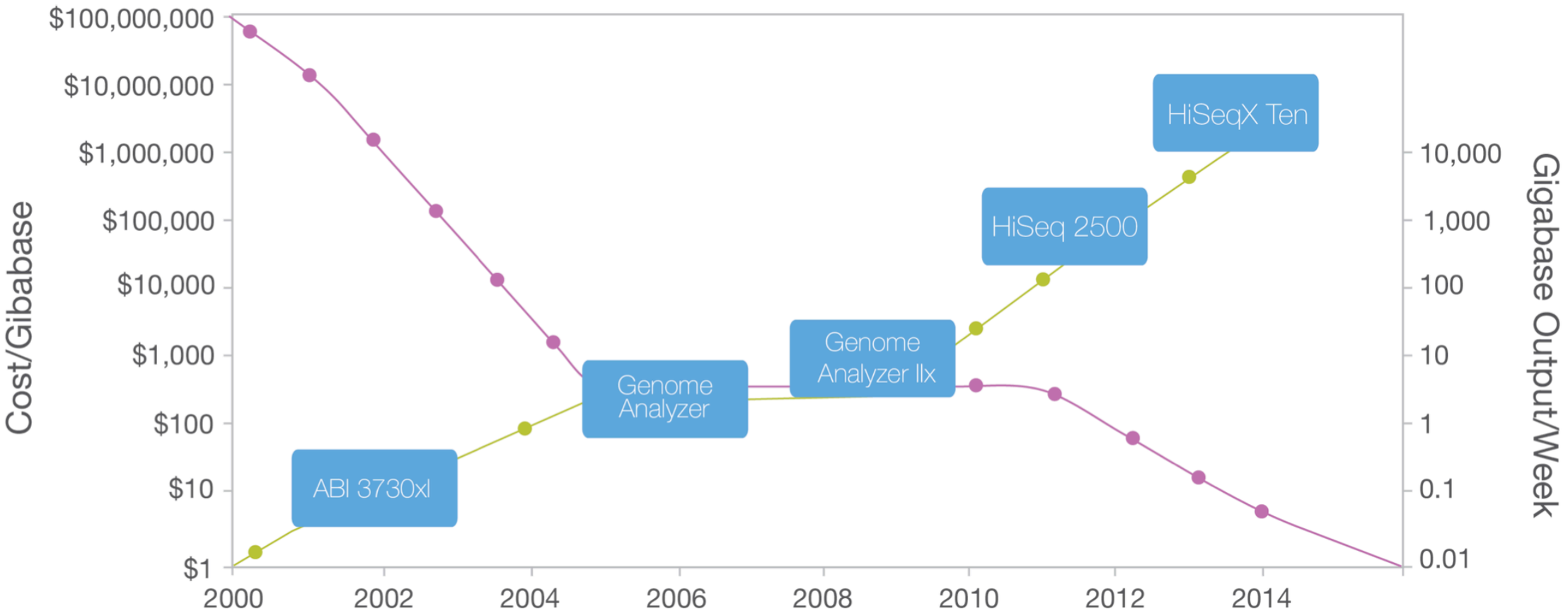


Sequencing by Synthesis (SBS)

- Sync through “reversible termination”.
- Line up all images; for each template/cluster, turn the colors into nucleotides.



High-throughput, Low Cost



- Cost of whole-genome sequencing: ~$1000.
- Promise of Illumina: $100 per individual in “very near future”.

3rd Generation Sequencing

1. Single-molecule sequencing (i.e., no PCR needed)
2. Direct sequencing (i.e., no synthesis)
3. Long reads: >200kb
4. High error rate: 5%-15% (HiFi reads: < 1%)
5. Representative technologies:
    1. Single-Molecule, Real-Time (SMRT), by PacBio
    2. Nanopore sequencing, by ONT

Official Videos

- PacBio
  - https://youtu.be/_lD8JyAbwEo
  - https://youtu.be/NHCJ8PtYCFc
- ONT
  - https://youtu.be/RcP85JHLmnI
  - https://youtu.be/E9-Rm5AoZGw