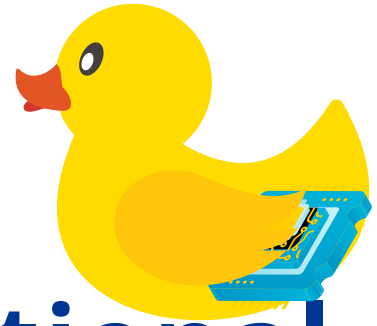


Computational Biology

(BIOSC 1540)

Lecture 02:

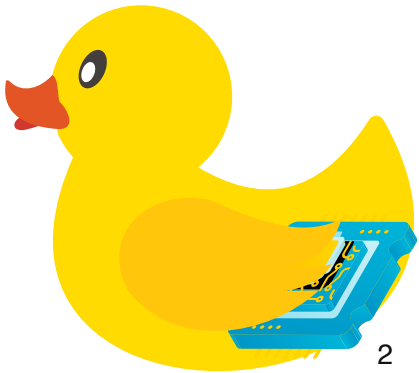
DNA sequencing

Aug 29, 2024

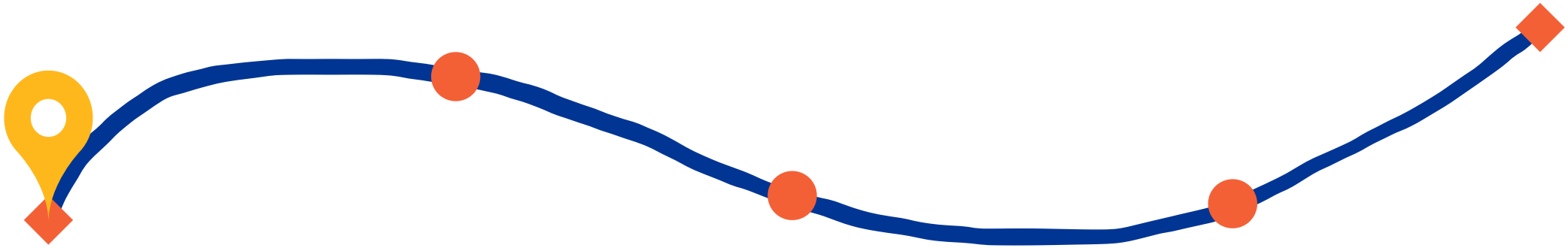
University of Pittsburgh

Announcements

- Assignment 01 will be published tonight or tomorrow.
- What material will you be responsible for
  - Anything covered on slides
  - Anything under the "Readings" subsection on the lecture page
- TopHat question about project.



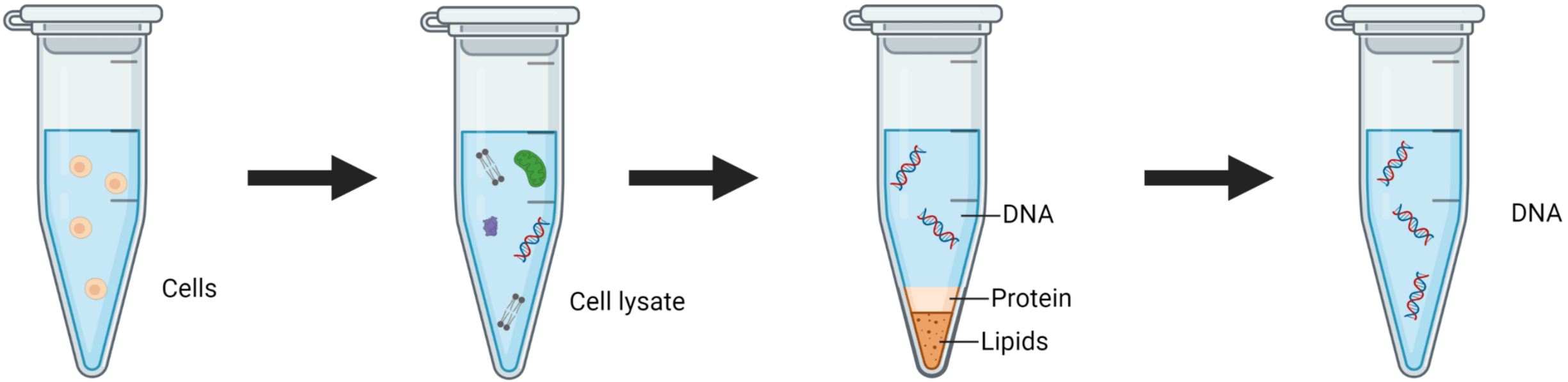
After today, you should be able to



1. **Construct a general workflow intrinsic to DNA sequencing experiments.**
2. Delineate the core principles underlying Sanger sequencing.
3. Conduct a comparative analysis of Illumina sequencing vis-à-vis Sanger sequencing.
4. Explicate the fundamental principles governing Nanopore sequencing technology.

How do we acquire our DNA sample?

Computationalists still need to understand the underlying source of our data



Let's start with a bacterial culture

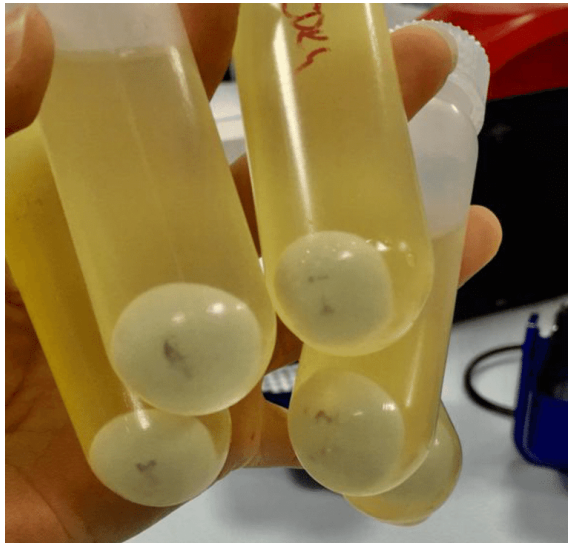
We let our bacterial culture produce our products of interest

Biotechnology frequently uses massive *E. coli* cultures to produce biologics

Fun fact: Pitt has a beer brewing class (ENGR 1933)

Separate cells from media

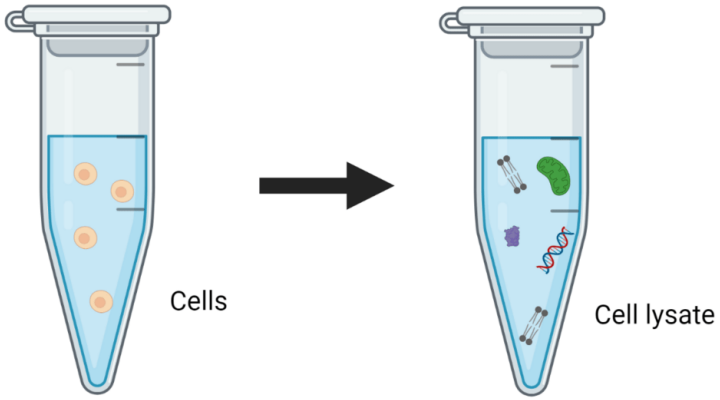
The first step is always to centrifuge and separate our cells and media



Keep the part that has our **component of interest** (DNA)

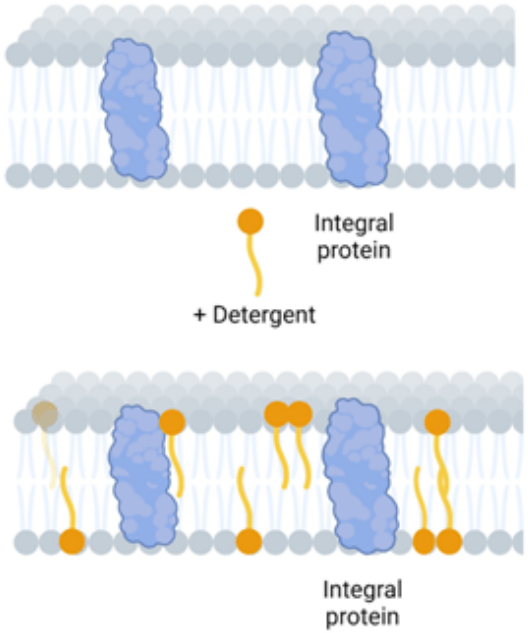
Great! We have our cells, but how can we get DNA out of our cells?

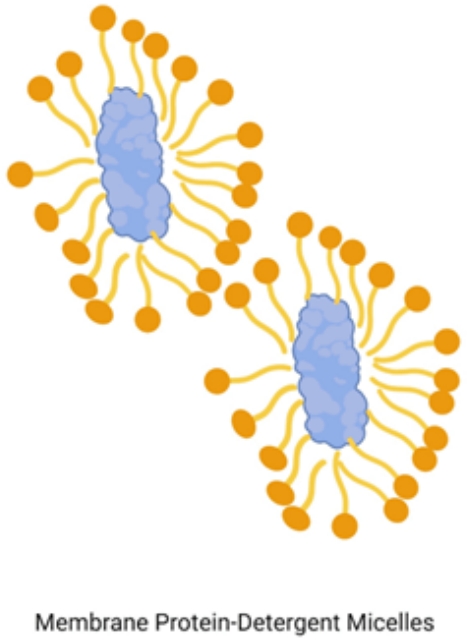
We **break open our cells** by lysing them



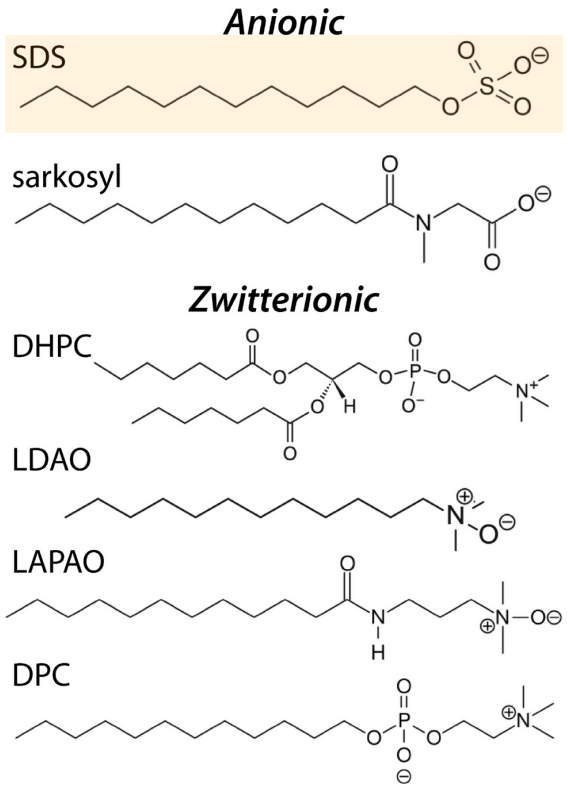
How can we lyse cells?

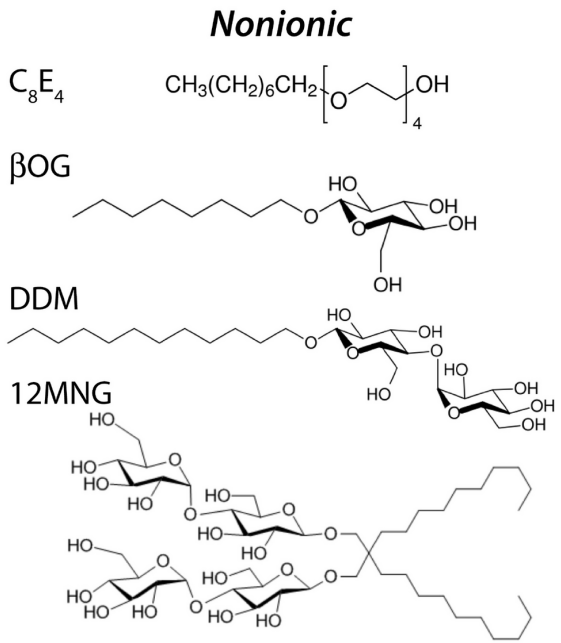
Chemical lysis destabilizes the lipid bilayer and denatures proteins

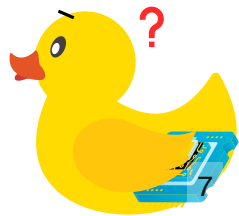




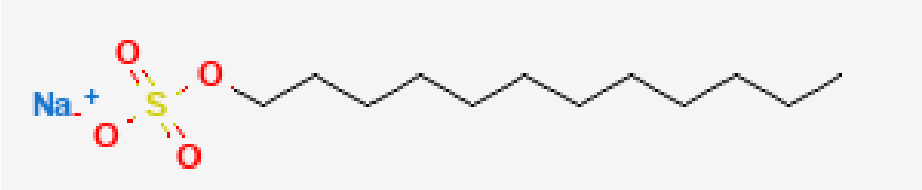
They have a hydrophilic head and hydrophobic tail



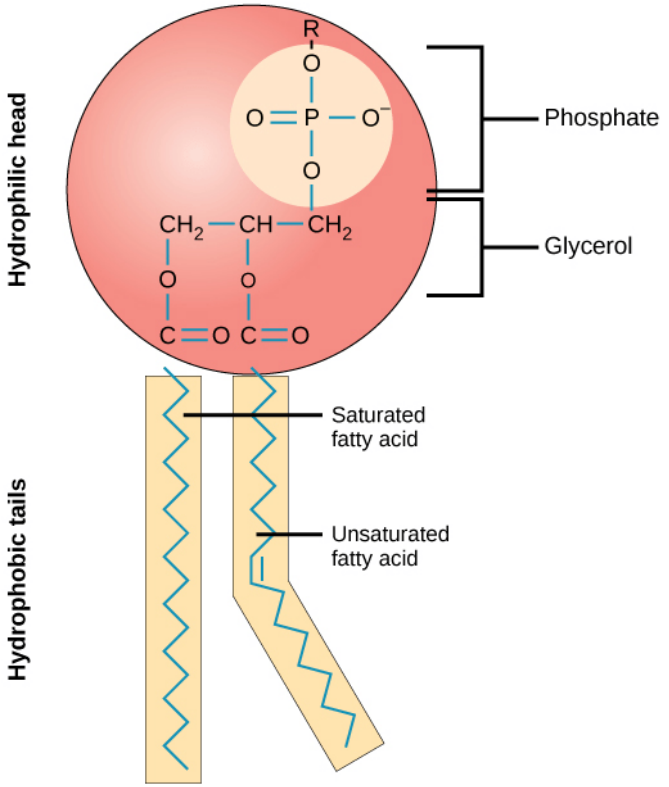




Wait, surfactants sound a lot like phospholipids?

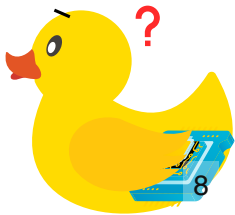


What's the primary difference, and how does this change its behavior?

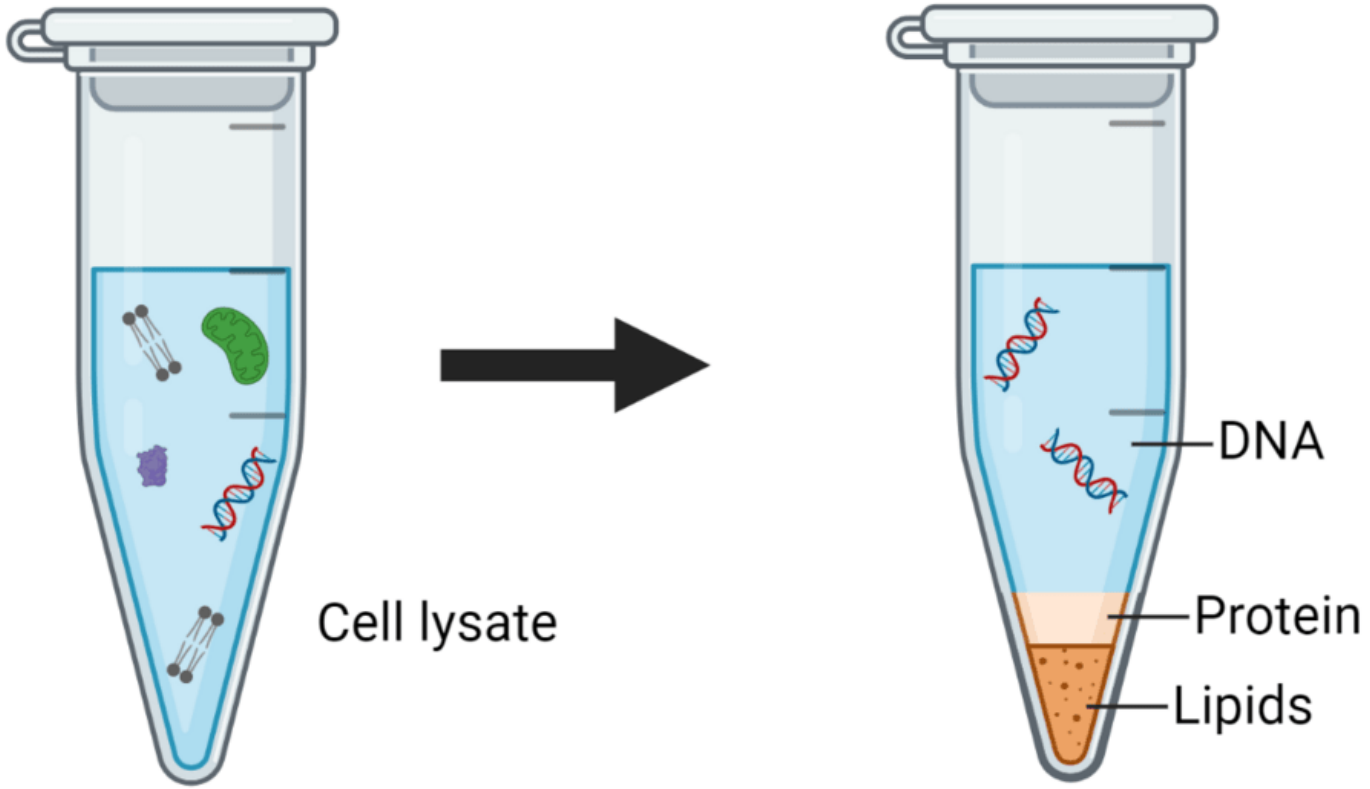


Surfactants have one **hydrophobic tail**, which allows them to further penetrate molecular structures

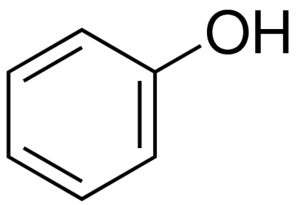
(There are also other methods like sonication.)



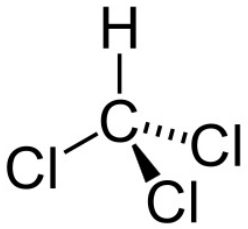
We need to isolate and purify our DNA



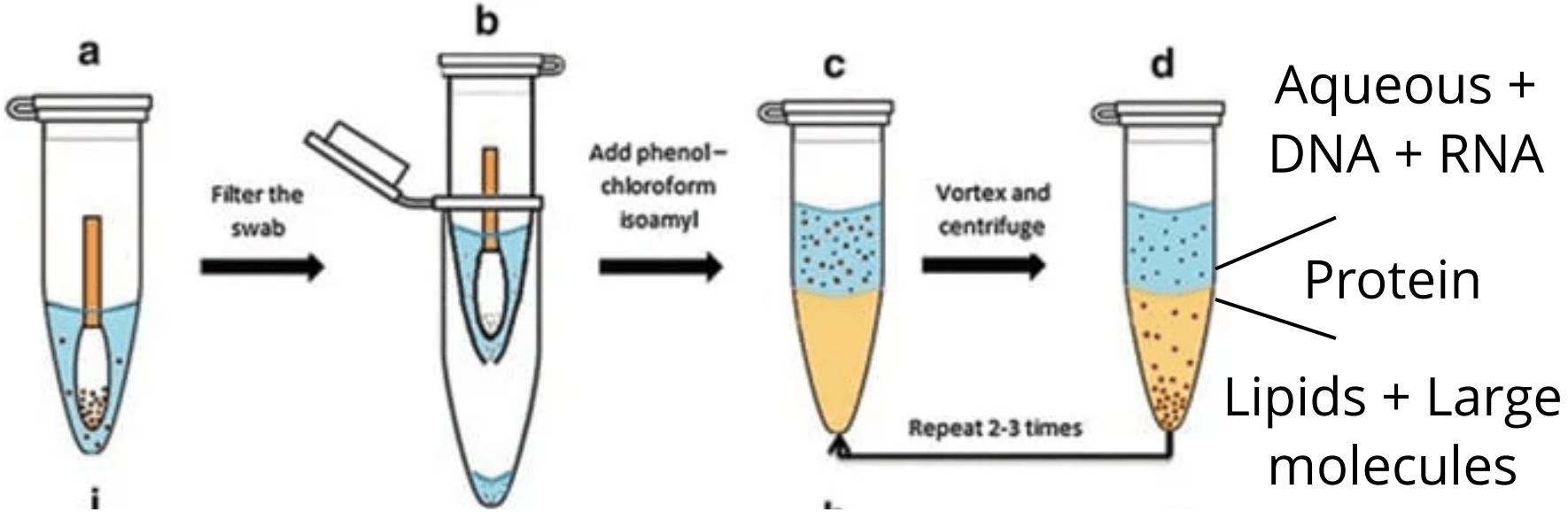
Phenol-chloroform extraction uses liquid-liquid separation



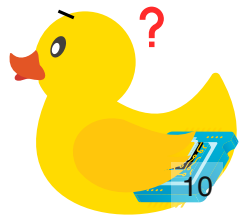
Phenol



Chloroform

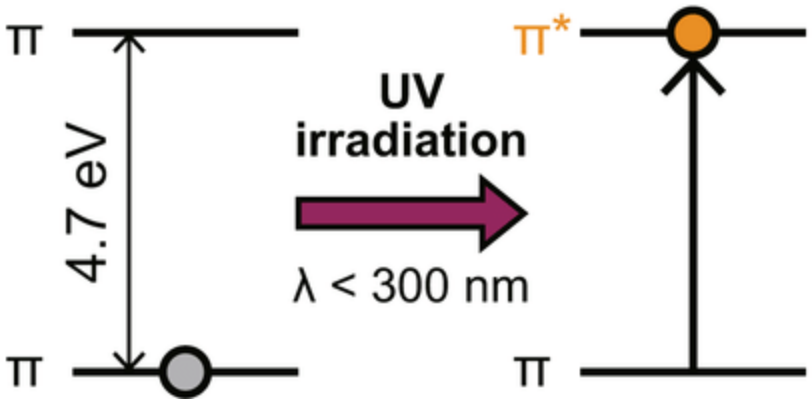


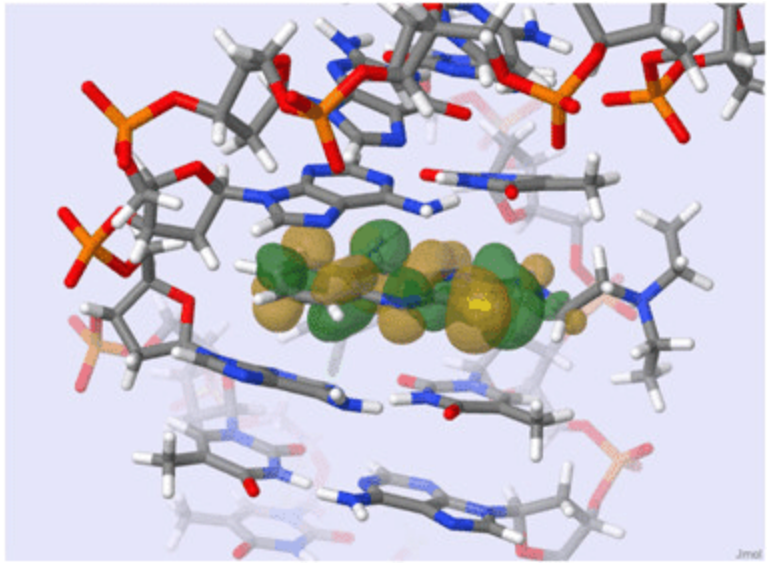
Where is our DNA, and why?
Which region should we keep?

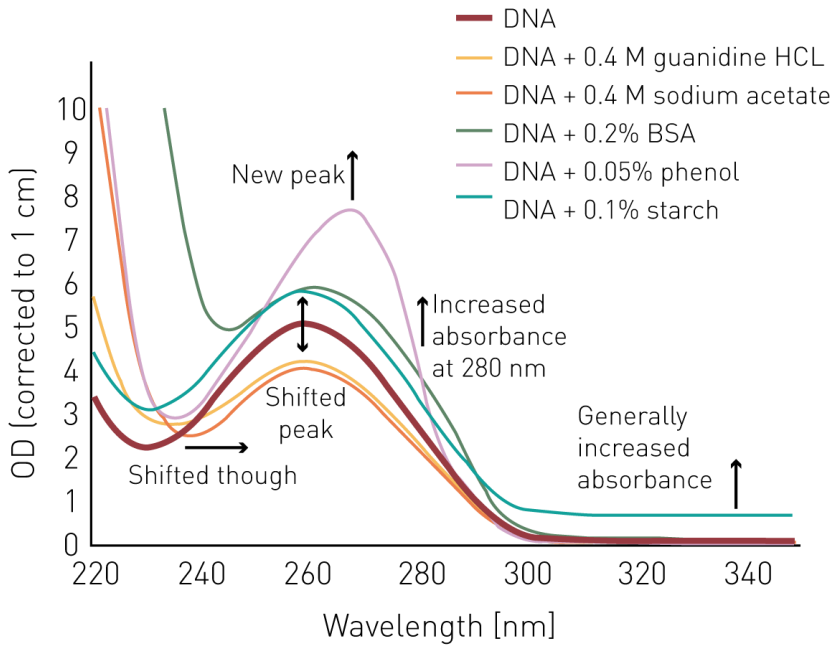


Most labs use highly effective kits

Sample absorbance at 260 nm is correlated to DNA concentration



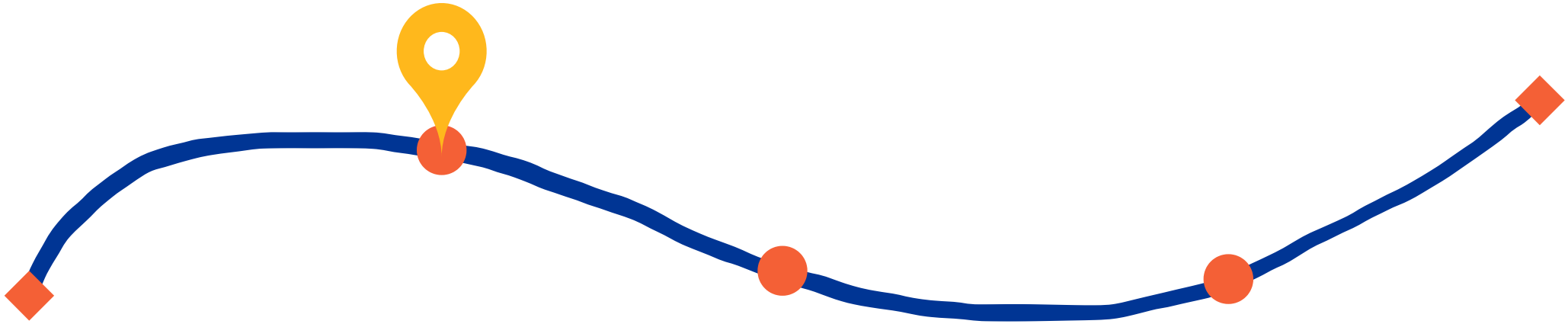




Substances	Ratio A260/A230	Ratio A260/A280	A340
Optimum from literature	*2.3-2.4*	*1.7-2.0*	
DNA	2.31	1.80	0.04
DNA + 0.4M guanidine HCL	1.48	1.74	0.04
DNA + 0.4M sodium acetate	1.12	1.76	0.04
DNA + 0.2% BSA	0.45	1.44	0.10
DNA + 0.05% phenol	2.00	1.59	0.01
DNA + 0.1% starch	2.12	1.67	0.66

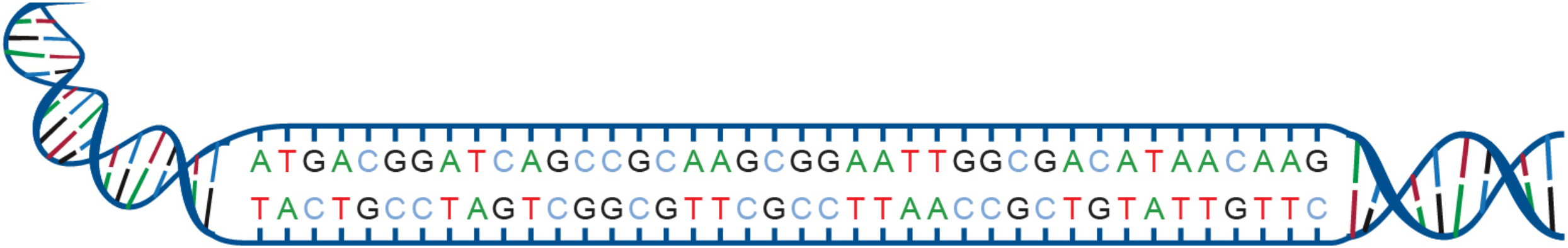
There are some other steps, but let's now assume we have a purified DNA sample at this point

After today, you should be able to



1. Construct a general workflow intrinsic to DNA sequencing experiments.
2. **Delineate the core principles underlying Sanger sequencing.**
3. Conduct a comparative analysis of Illumina sequencing vis-à-vis Sanger sequencing.
4. Explicate the fundamental principles governing Nanopore sequencing technology.

Our main problem: Determine the precise ordering of nucleotides

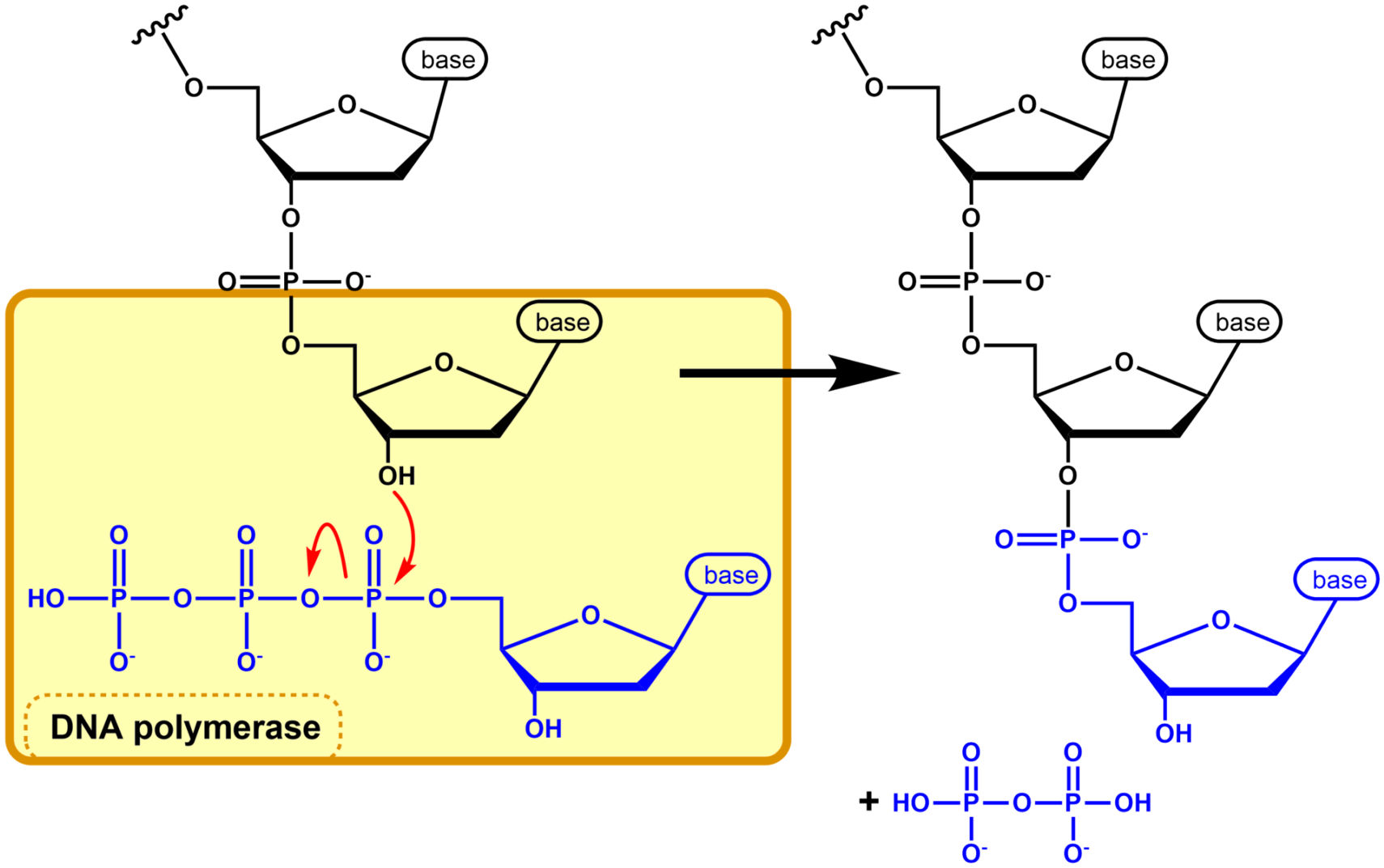


DNA elongation happens rapidly and continuously

We use DNA polymerase + excess nucleotides to make copies of DNA

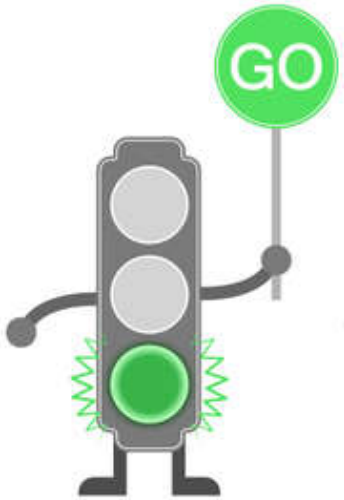
https://omics.crumblearn.org/sequencing/dna/pcr/dna-elongation.html

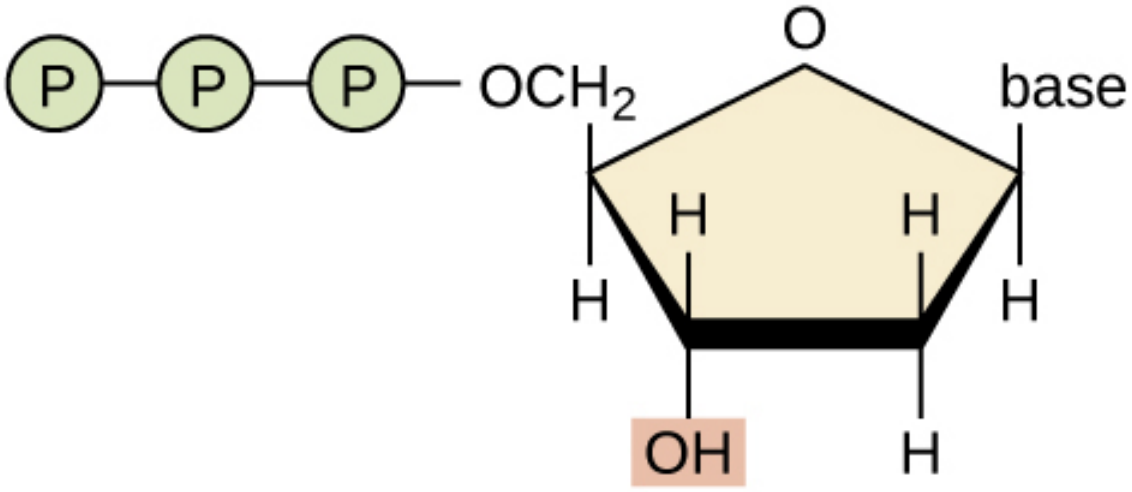
3' OH is required for DNA elongation



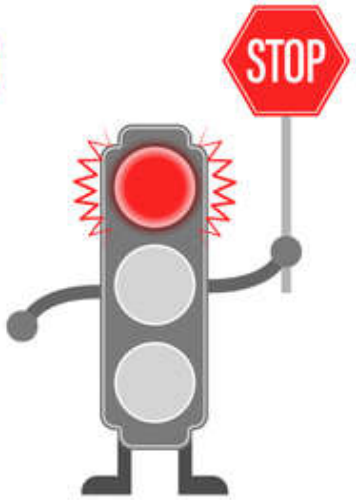
What happens if we don't have the 3' OH? We cannot add another nucleotide

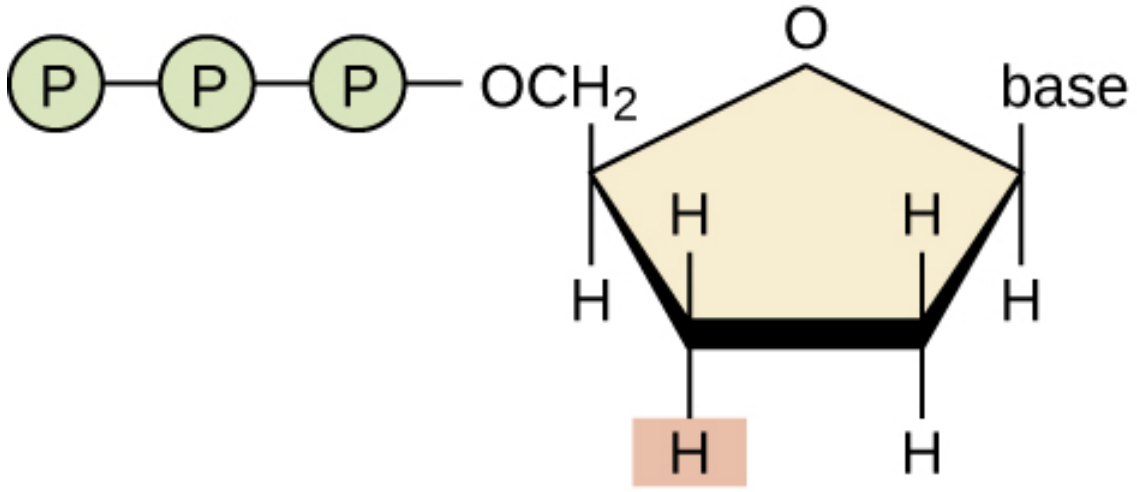
Di-deoxynucleotides stop replication





deoxynucleotide (dNTP)





dideoxynucleotide (ddNTP)

ddNTP will randomly stop DNA elongation

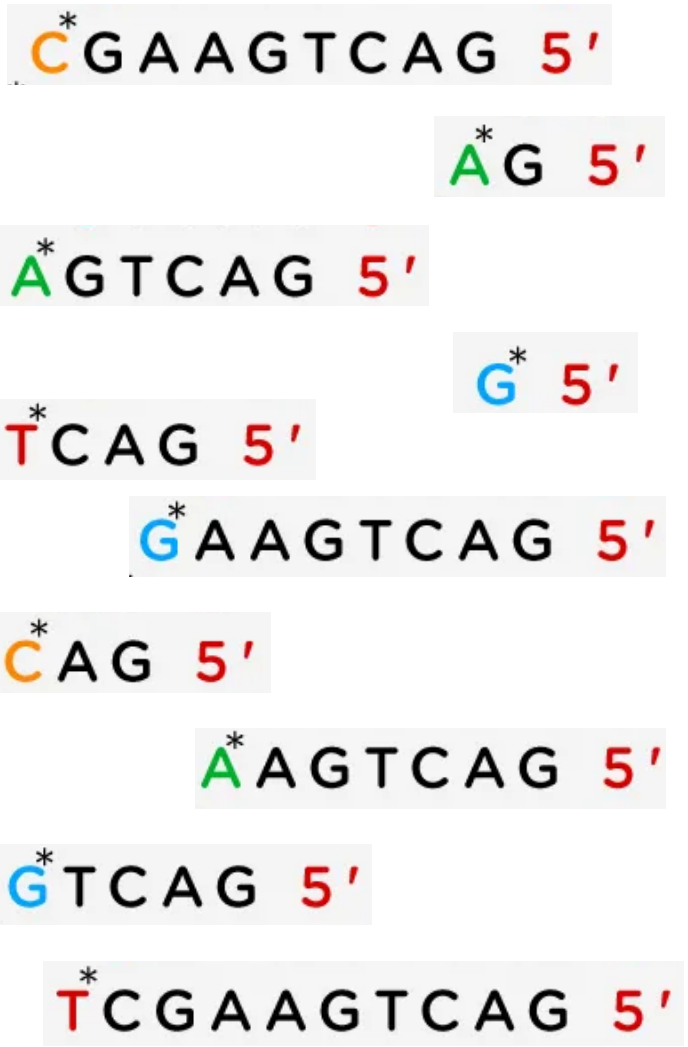
When DNA polymerase adds a **ddNTP**, it cannot add any other nucleotide

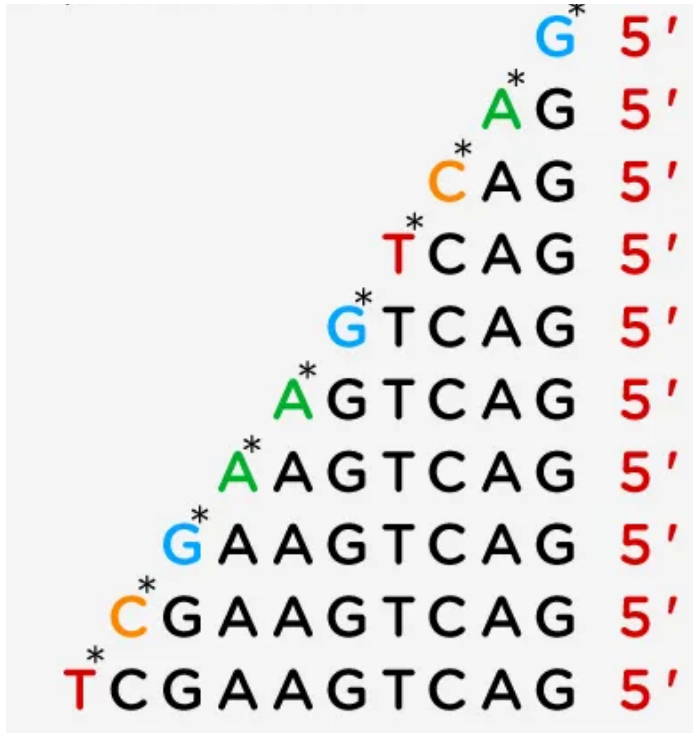
Ratio is usually **1**:**100**

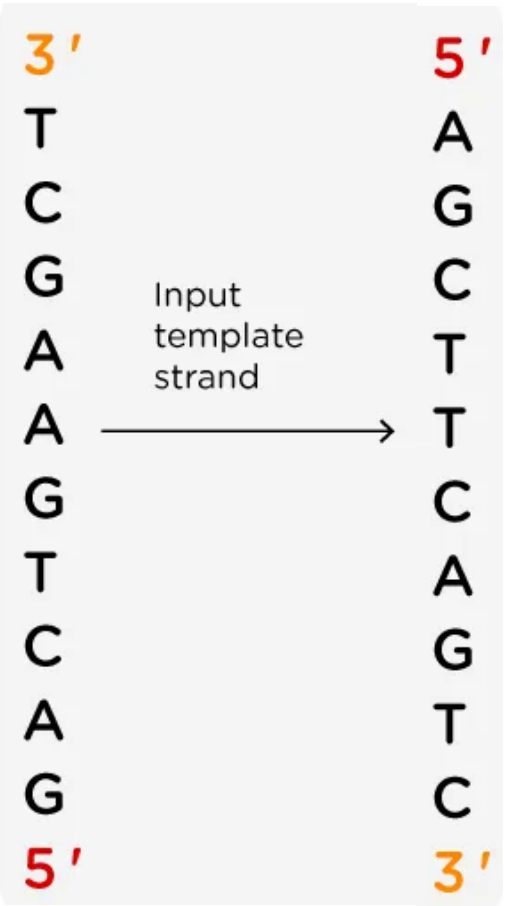
We will be left with DNA strands of variable length

https://omics.crumblearn.org/sequencing/dna/first-gen/sanger/principles/chain-termination.html

By sorting DNA fragments by length, we can see what the last nucleotide is







Original setup

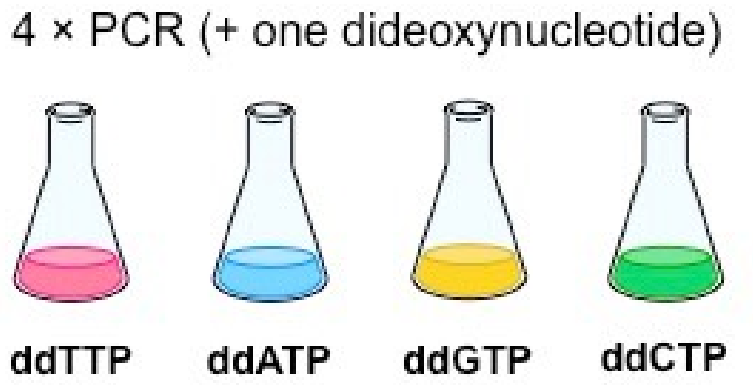
1. Split DNA sample into four beakers
2. Add all four dNTPs to each beaker
3. Add some amount of radioactive ddNTP in a single beaker
4. Add Taq polymerase and let PCR run

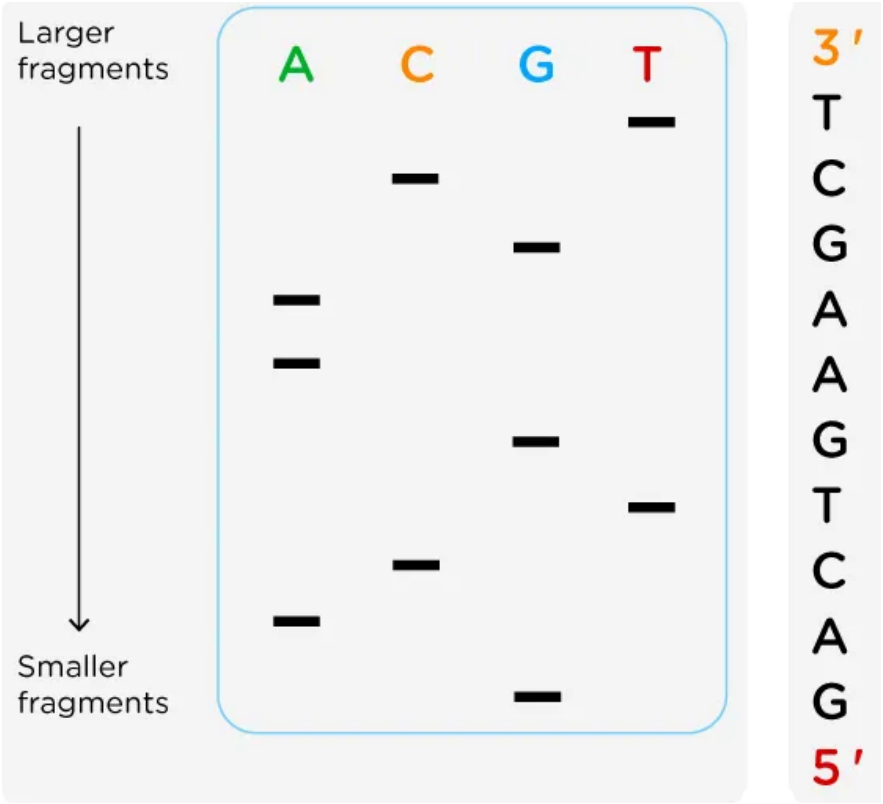
Once we have fragments, how can we separate them by length?

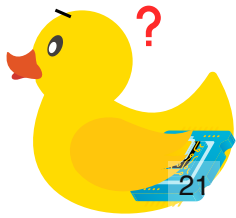
Gel electrophoresis!

Why would we need separate beakers?

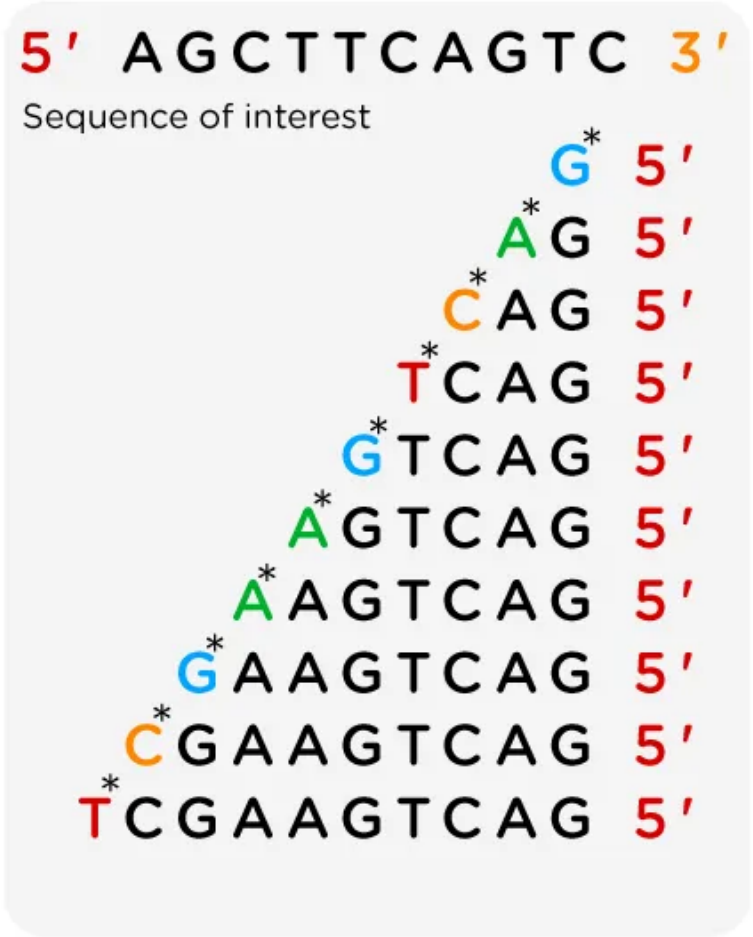
Cannot differentiate between radioactive nucleotides

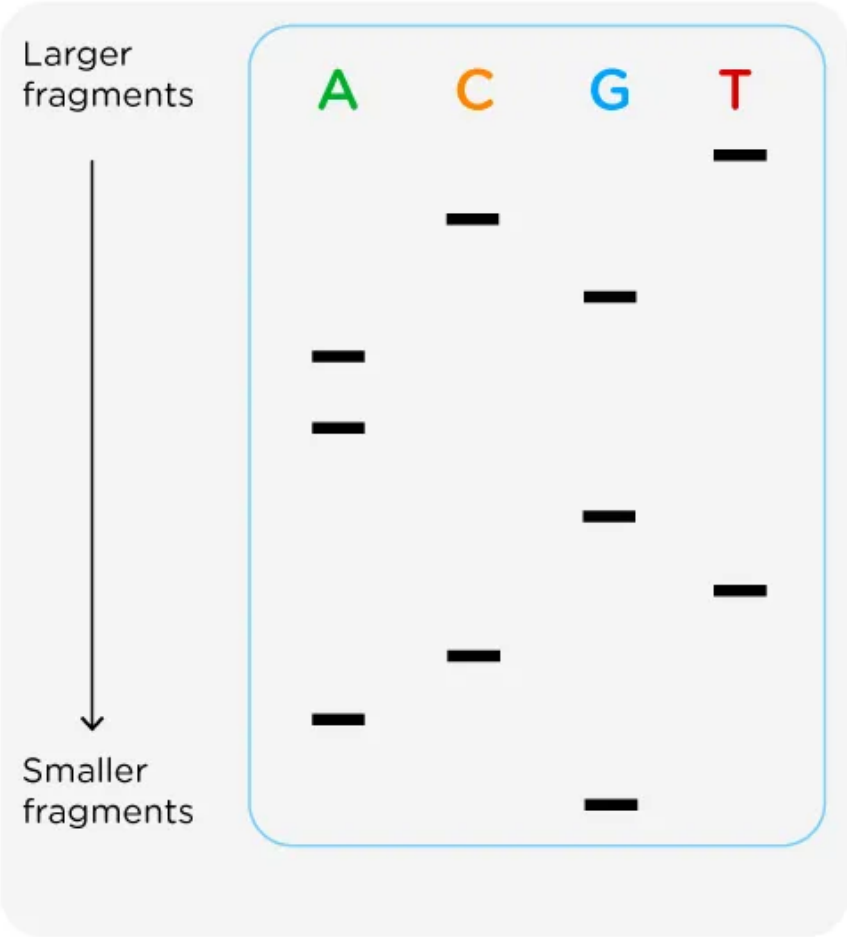


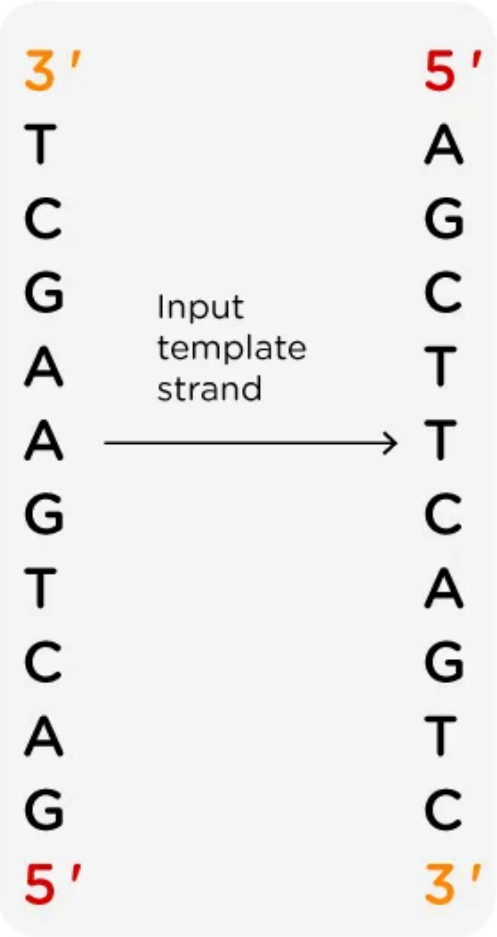




We can build our sequence based on what (radioactive) ddNTP is at that position

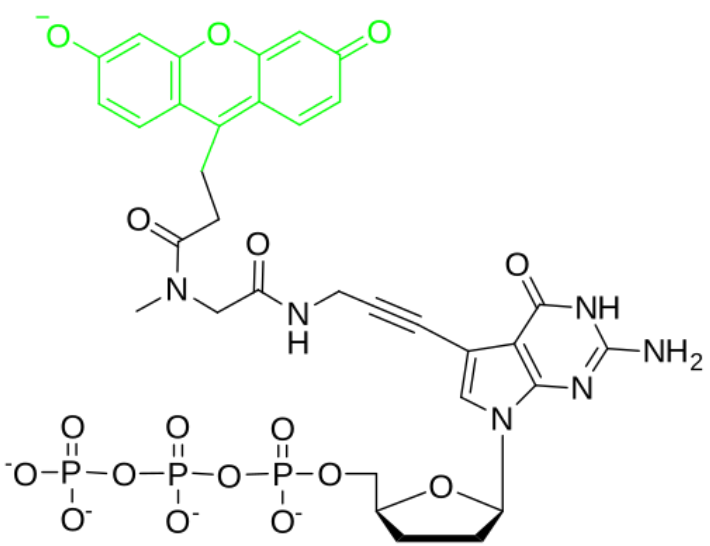




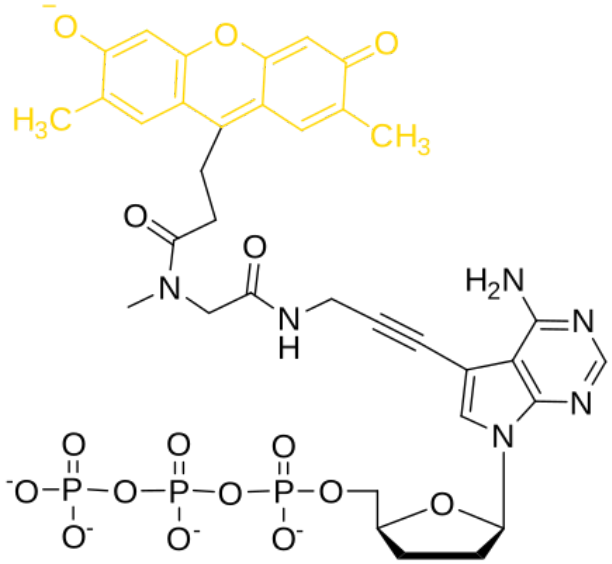


Now we use fluorescence to distinguish ddNTPs

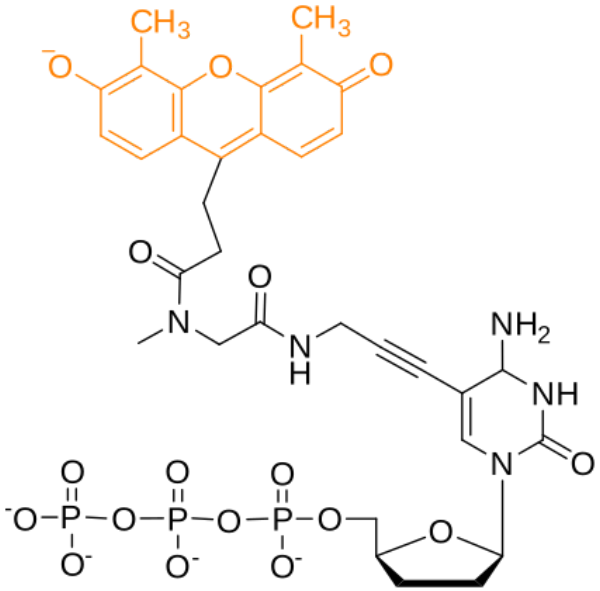
Only need one PCR!



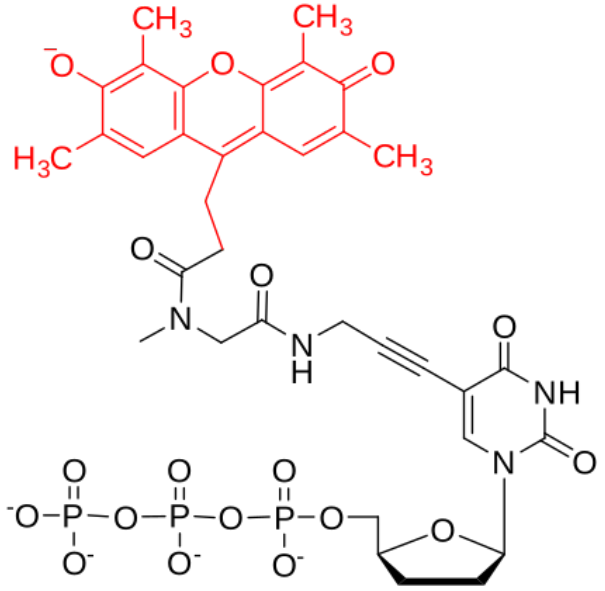
G-505



A-512



C-519

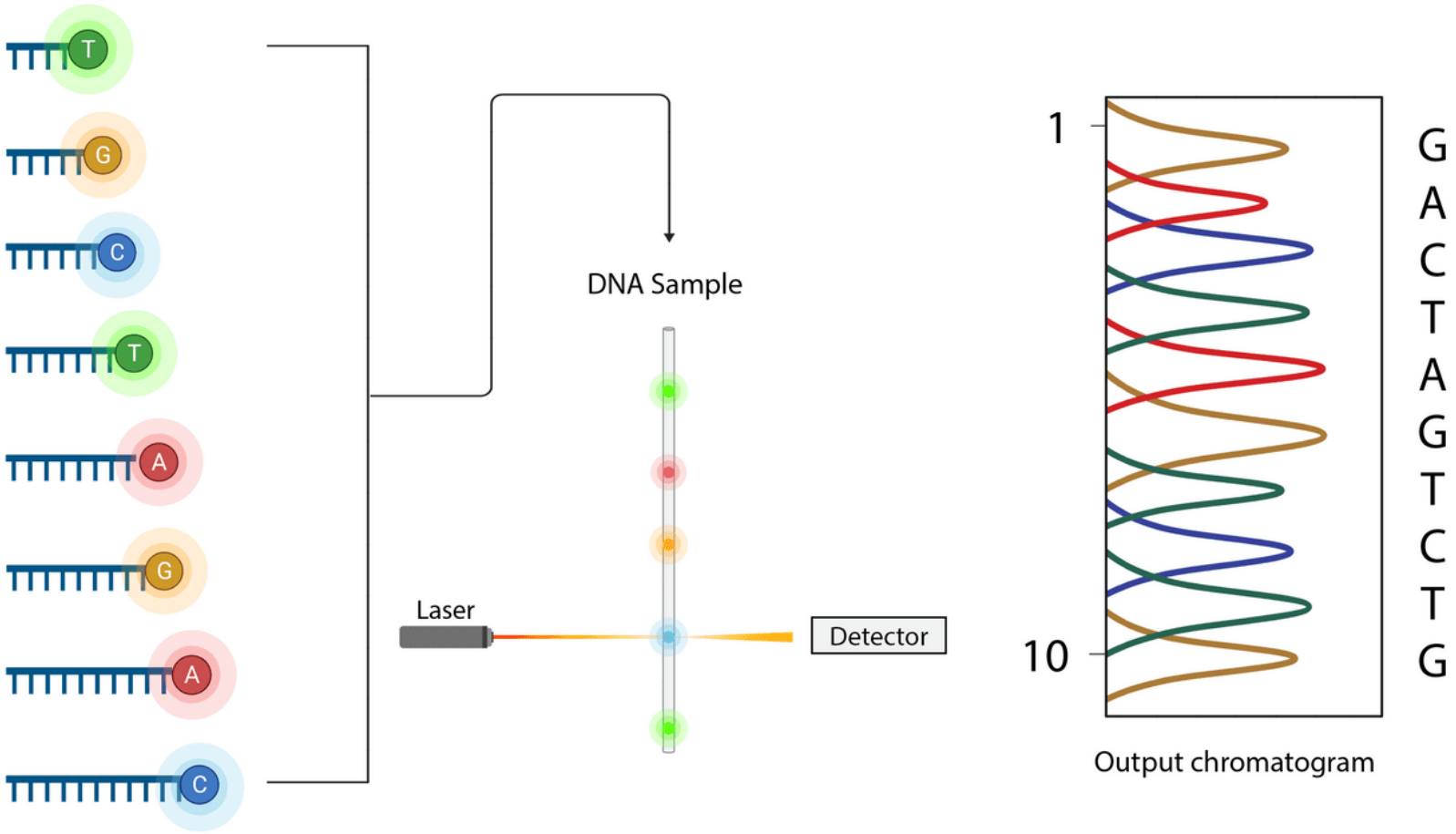


T-526

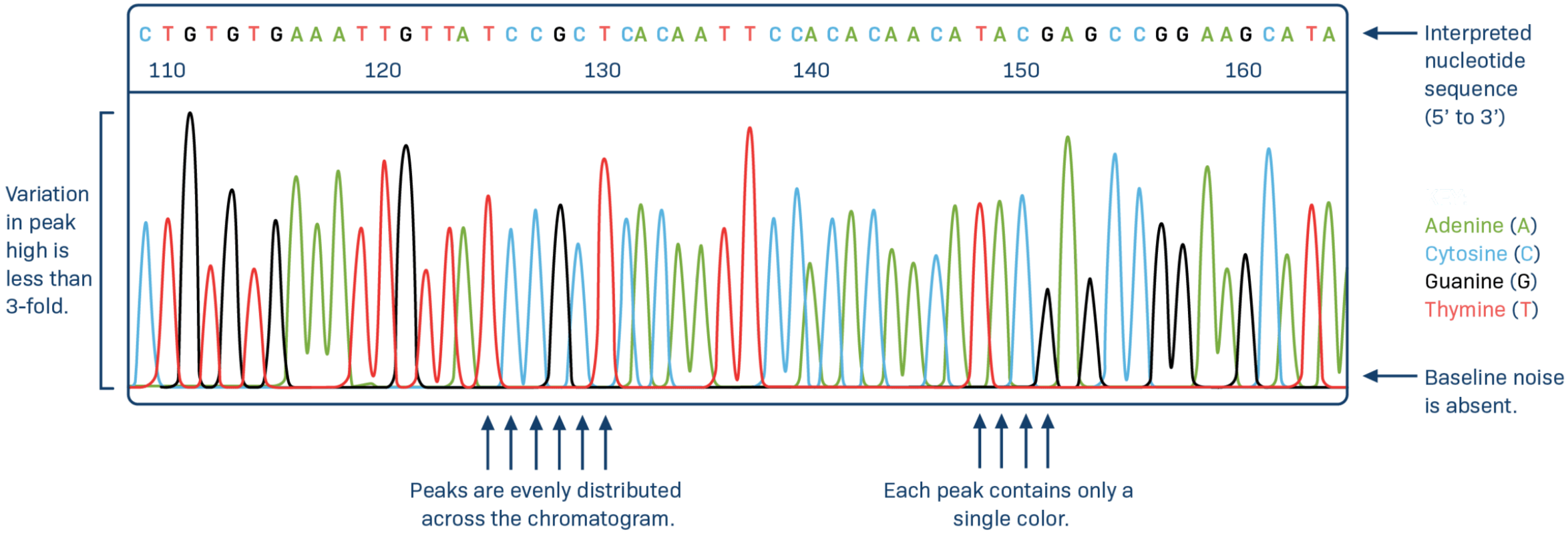
We also can automate fragment separation

Capillary gel electrophoresis can accelerate fragment length sorting and detection

Unique fluorescence signal per ddNTP produces a **chromatogram**

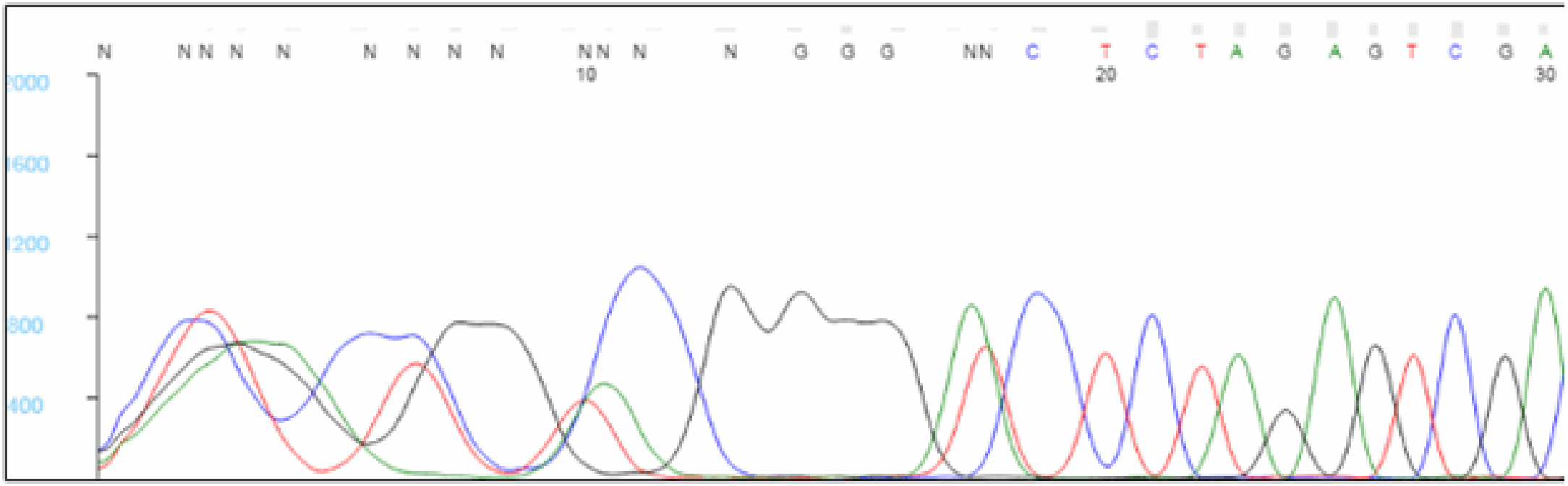


Ideal chromatogram

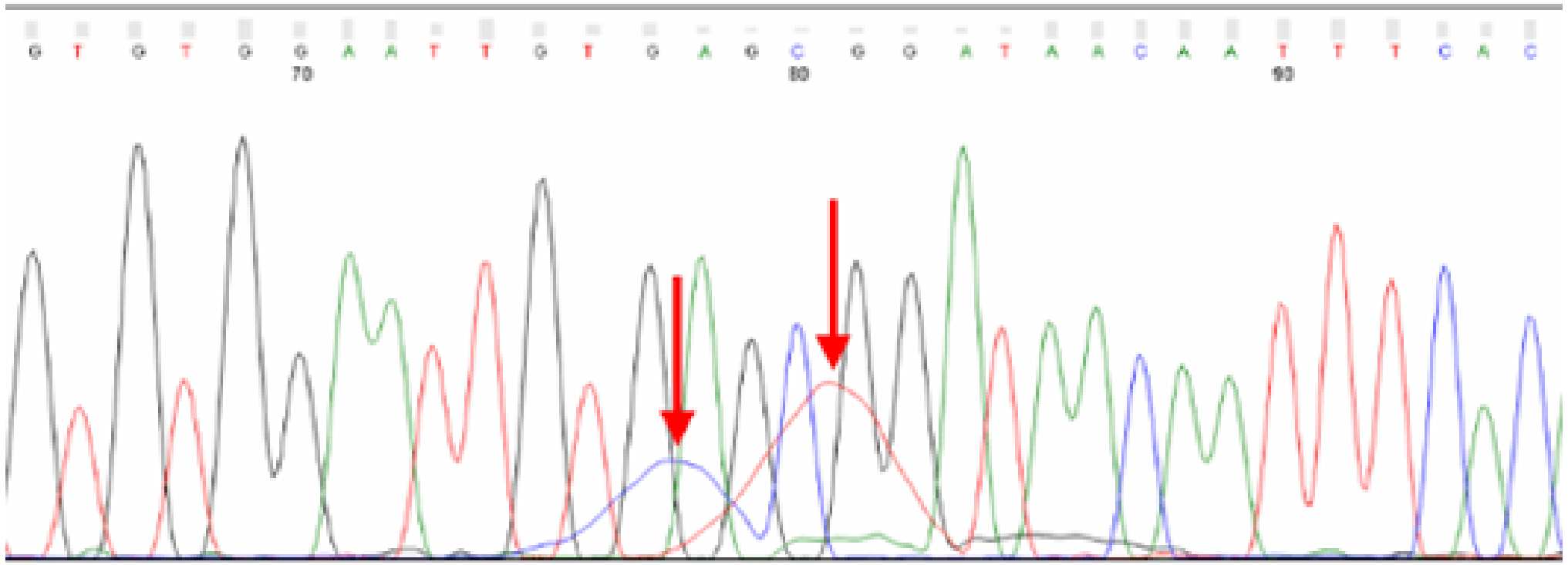


Significant noise up to ~20 basepairs in

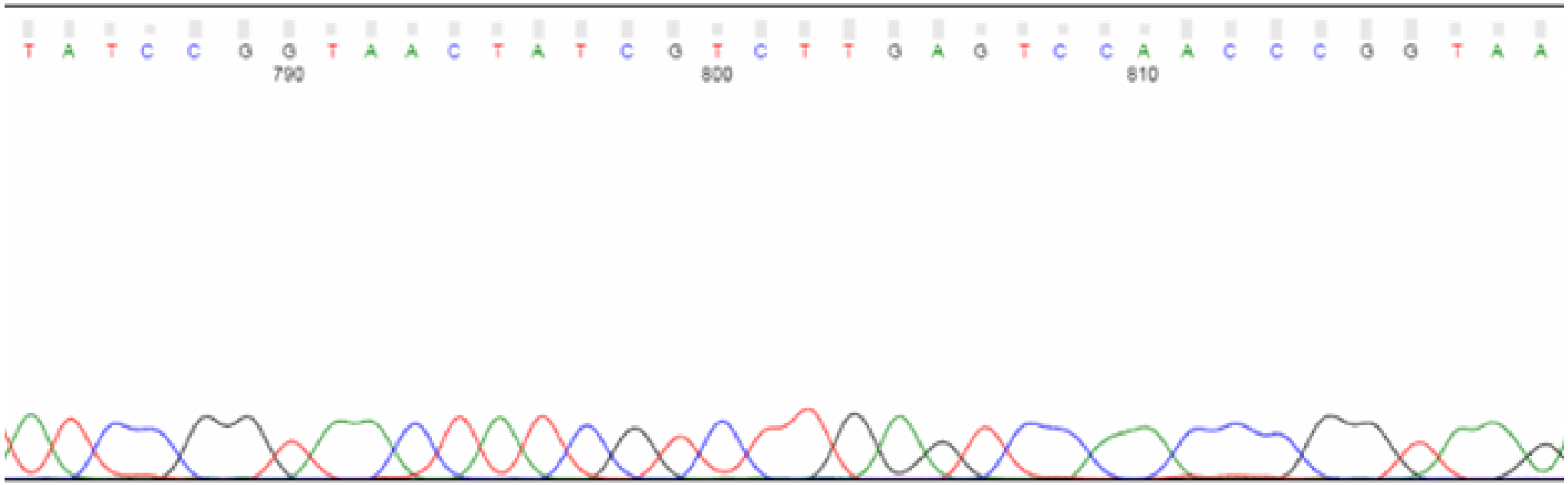
Unreliable transport properties



Dye blobs occur from unused ddNTPs



We have fewer longer fragments so signal is weaker

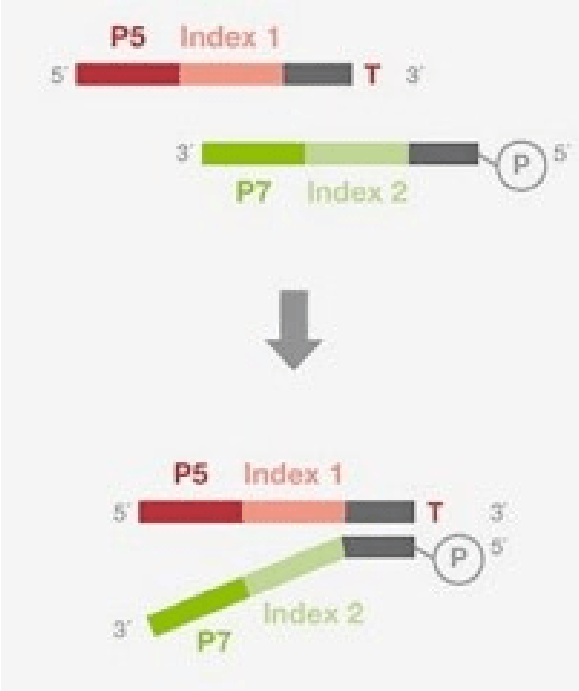


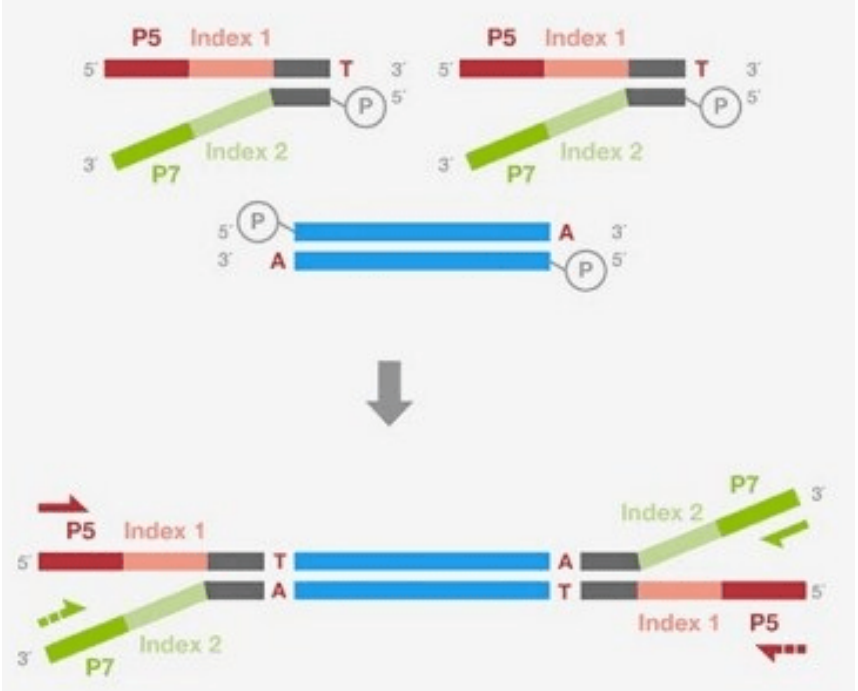
After today, you should be able to

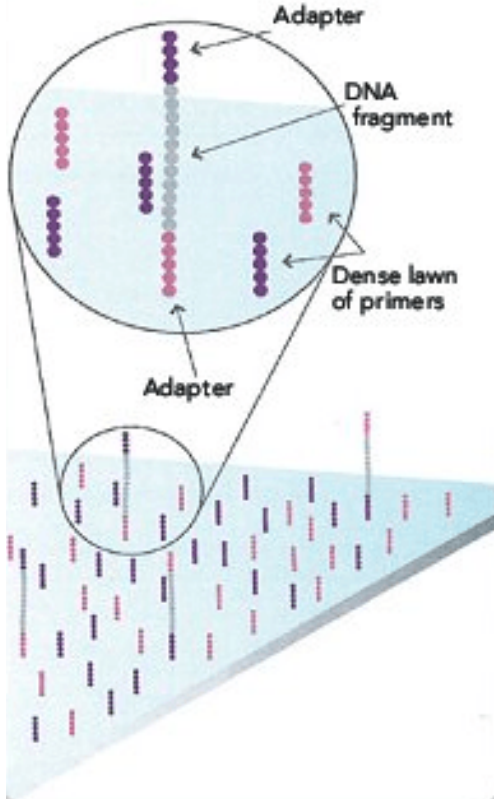
1. Construct a general workflow intrinsic to DNA sequencing experiments.
2. Delineate the core principles underlying Sanger sequencing.
3. **Conduct a comparative analysis of Illumina sequencing vis-à-vis Sanger sequencing.**
4. Explicate the fundamental principles governing Nanopore sequencing technology.

What is better than promotional materials?

Adapter ligations attach P5 and P7 oligos to facilitate binding to flow cell (Illumina)



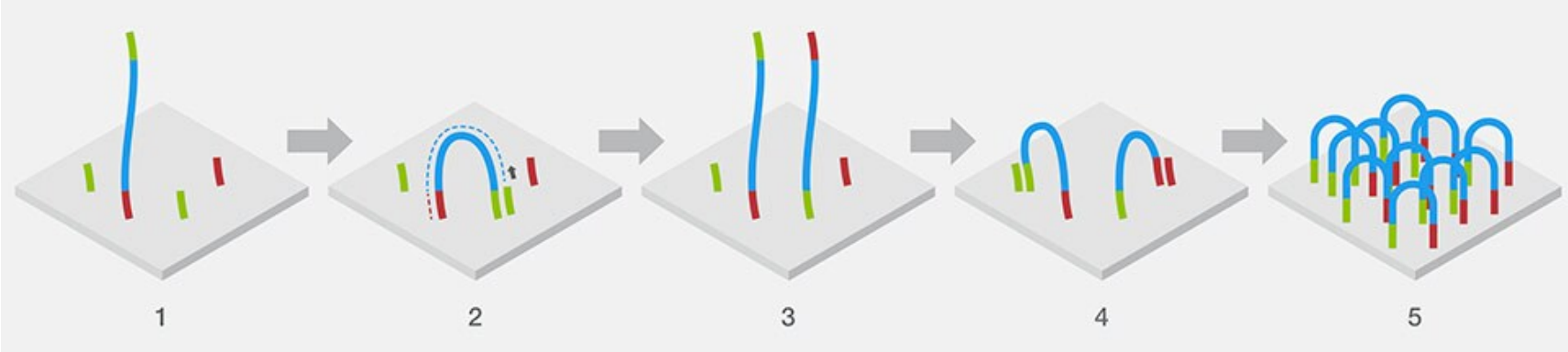




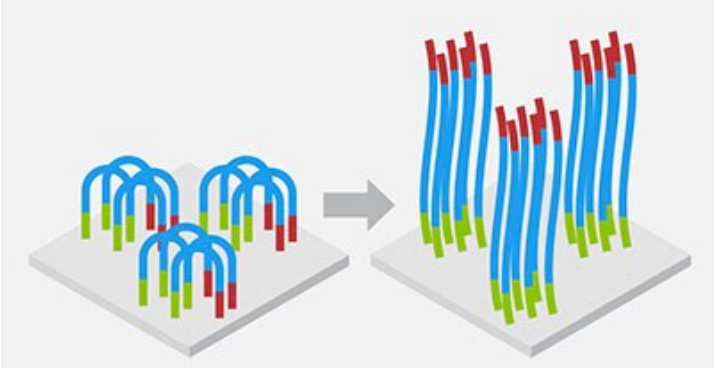
Primers are not complementary, so they do not base pair

We locally amplify bound DNA fragments to get clusters of the same sequence

Bridge amplification creates double-stranded bridges



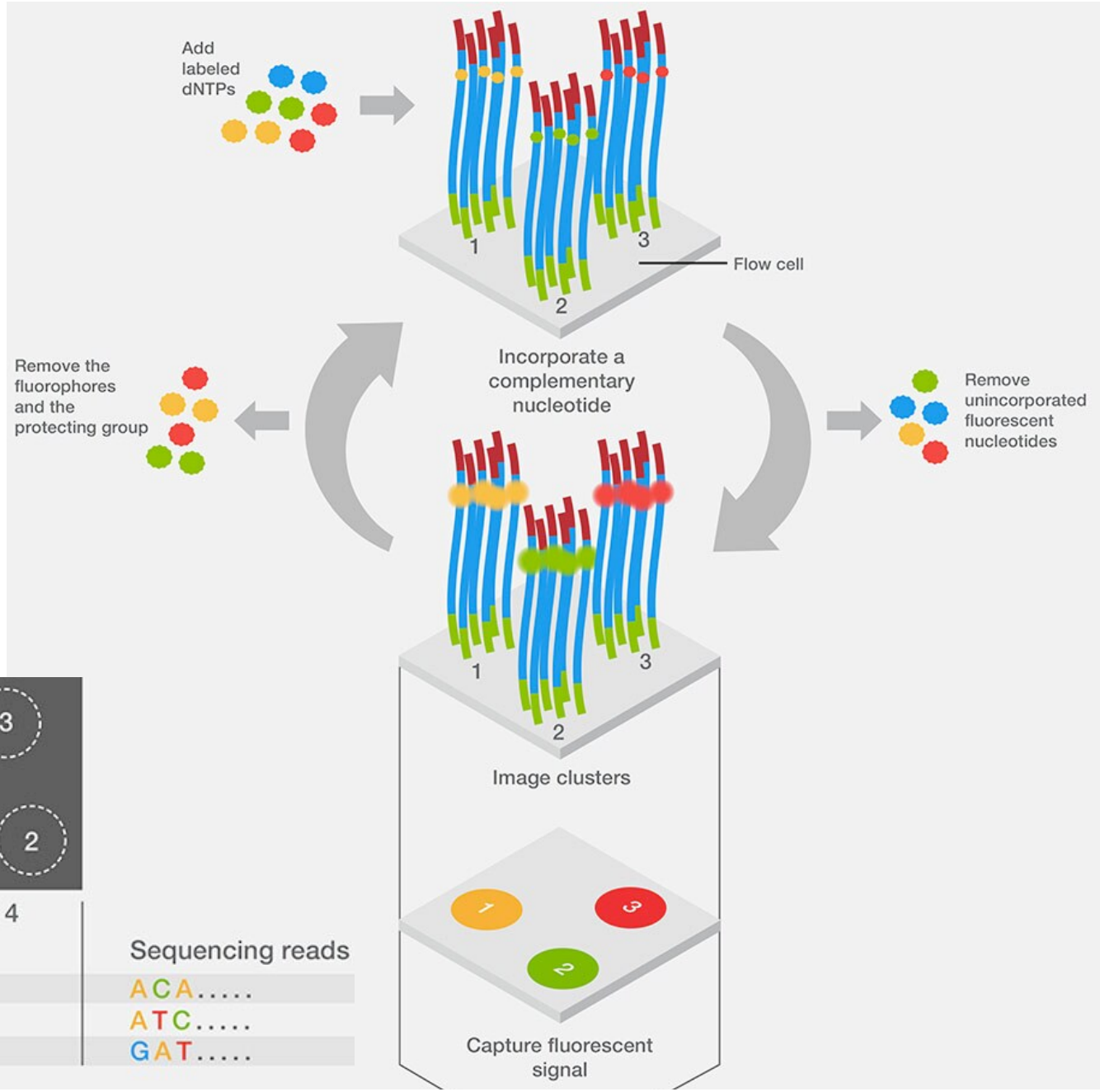
Double-stranded clonal bridges are denatured with cleaved reverse strands



Clusters will give off a stronger signal compared to a single fragment

We repeatedly

- Add nucleotide
- Capture signal
- Cleave fluorophore



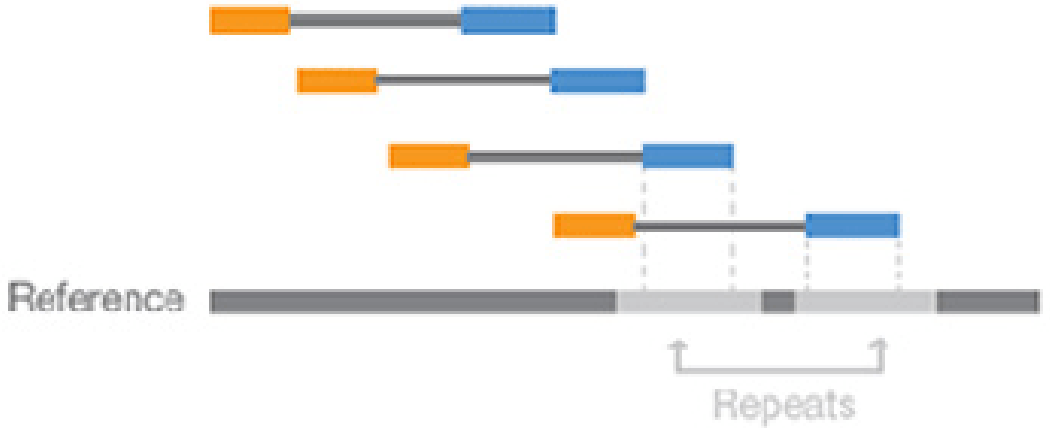
Calls	Cycle 1	Cycle 2	Cycle 3	Cycle 4	Sequencing reads
1	A	C	A	–	ACA.....
2	A	T	C	–	ATC.....
3	G	A	T	–	GAT.....

Paired-End Reads

Alignment to the Reference Sequence

Forward

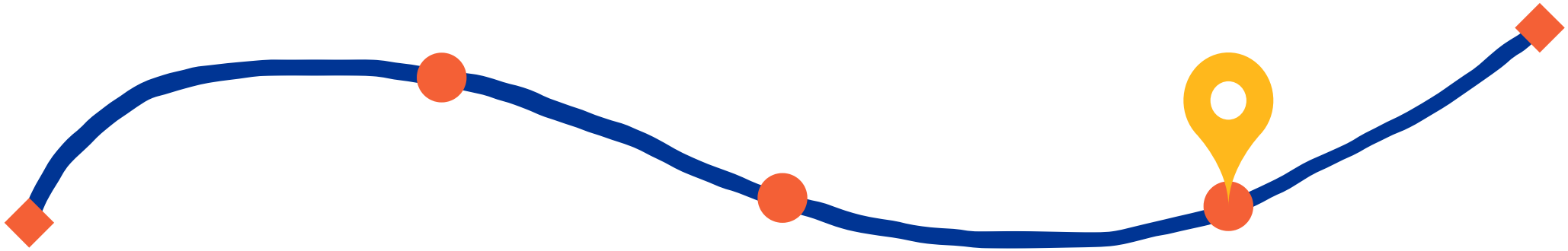
Reverse



Paired-end sequencing enables both ends of the DNA fragment to be sequenced. Because the distance between each paired read is known, alignment algorithms can use this information to map the reads over repetitive regions more precisely. This results in much better alignment of the reads, especially across difficult-to-sequence, repetitive regions of the genome.

Illumina is high throughput and widely used

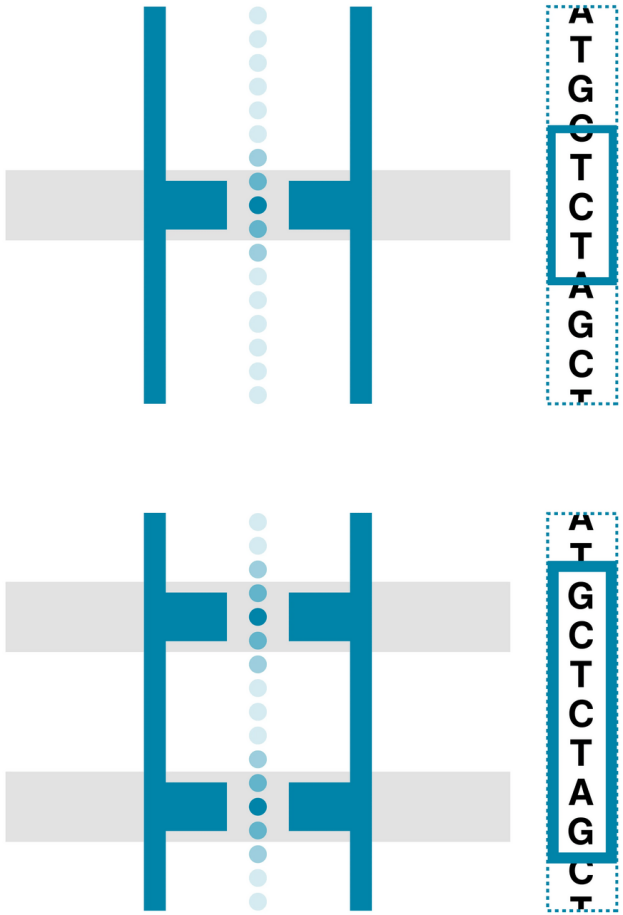
After today, you should be able to

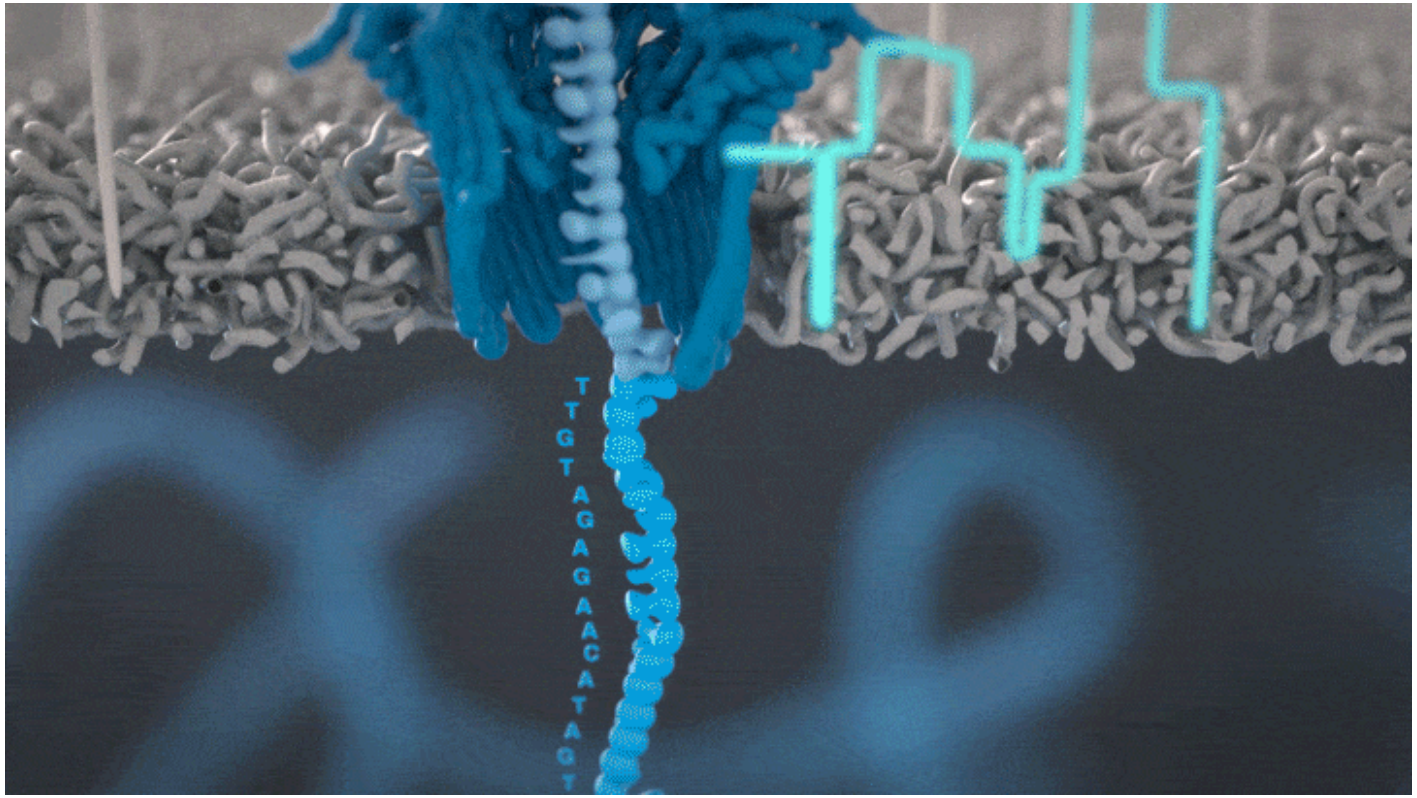


1. Construct a general workflow intrinsic to DNA sequencing experiments.
2. Delineate the core principles underlying Sanger sequencing.
3. Conduct a comparative analysis of Illumina sequencing vis-à-vis Sanger sequencing.
4. **Explicate the fundamental principles governing Nanopore sequencing technology.**

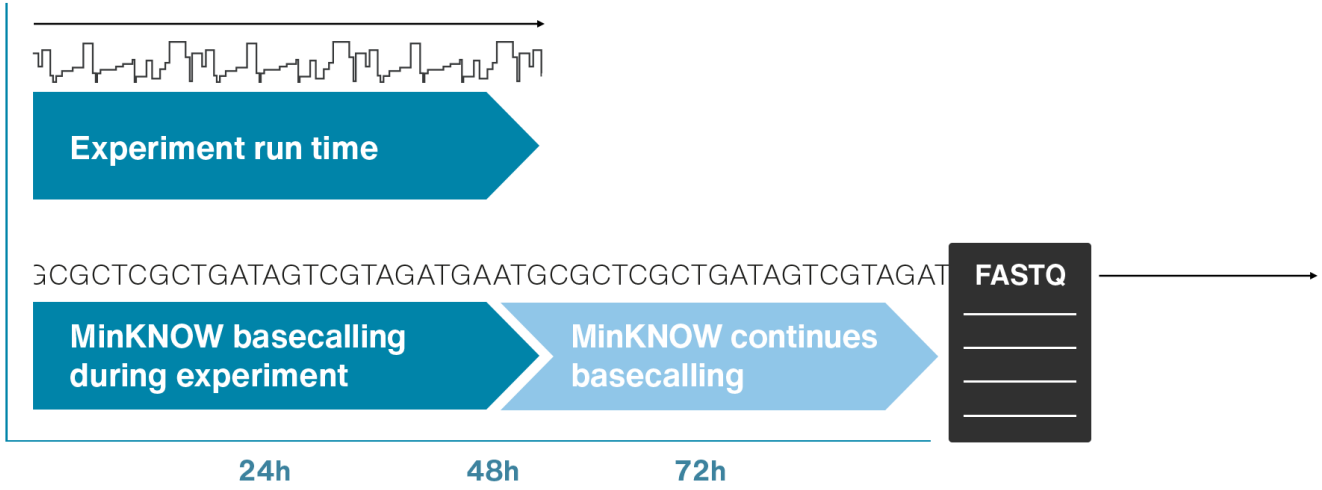
What is better than promotional materials?

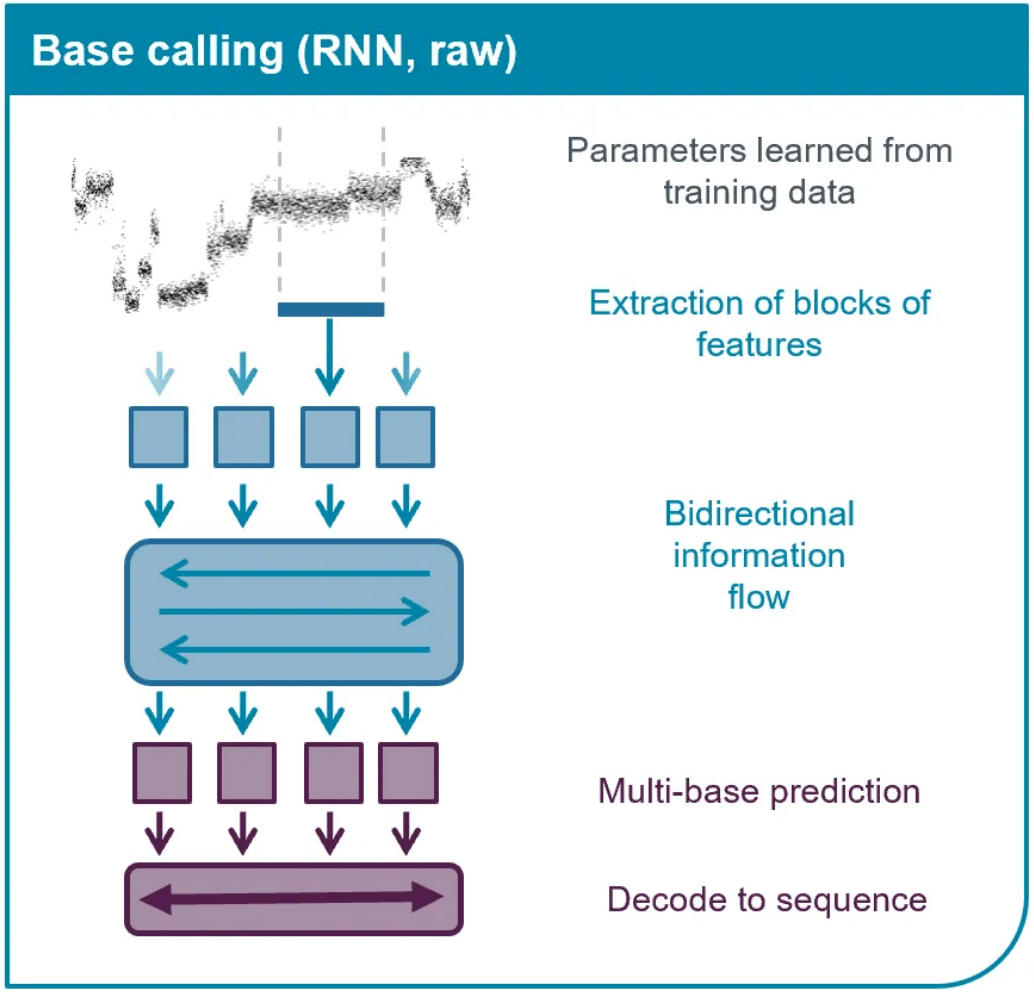
Nanopores and polymer membrane respond to electrical perturbations





ML algorithms predict and decode sequences

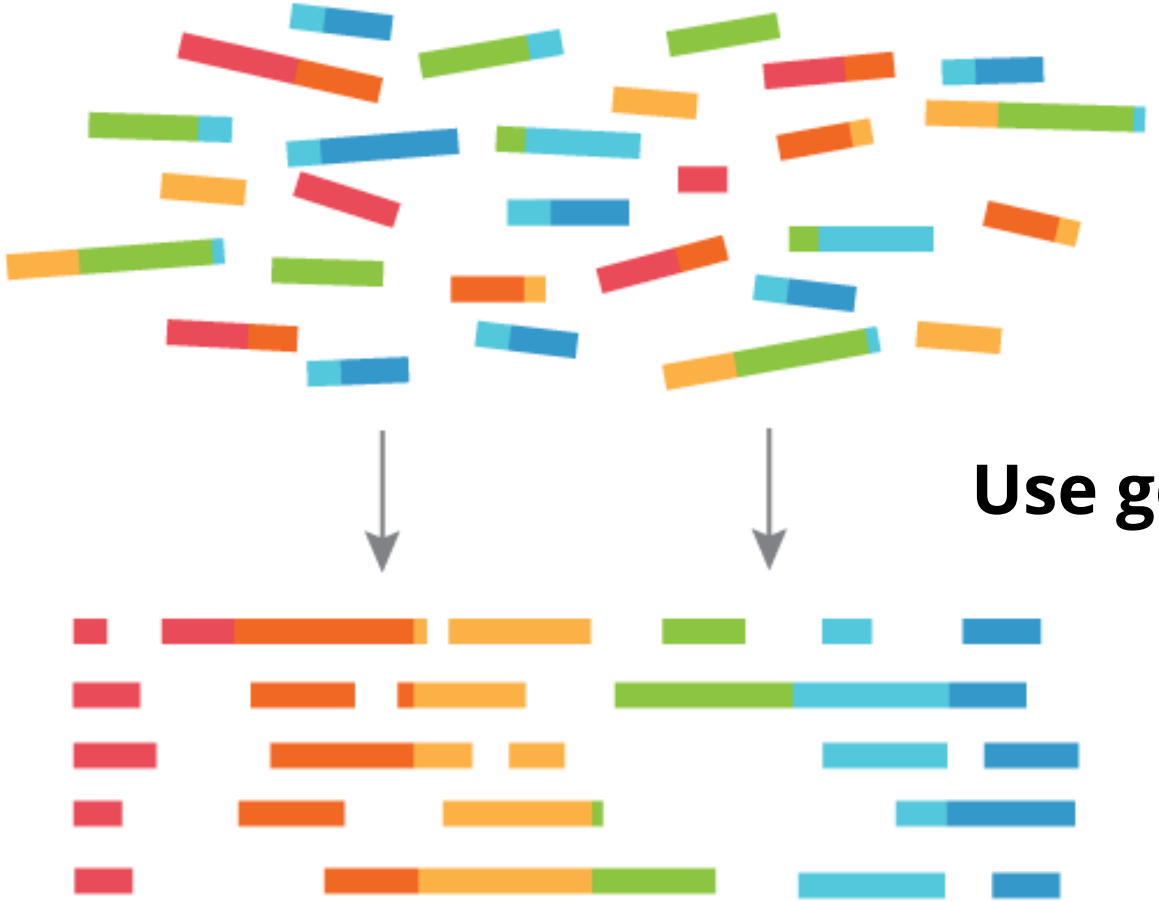




Nanopore gives us much longer reads, which is important for assembling reads into a genome

Sneak peek of the next lecture …

What we sequence



Use genome assembly!

What we want

ATGTTCCGATTAGGAAACCTATCTGTAACTGTTTCATTCAGTAAAAGGAGGAAA

Before the next class, you should

Lecture 02:
DNA sequencing

Lecture 03:
Sequencing quality control

Today

Tuesday

- Start Assignment 01, which will be released tomorrow.