Lecture 6:

DNA Sequencing

The thing to the information

Course 485

Introduction to Genomics

AIMS

- Introduce DNA sequencing technologies.
- Highlight method of sequencing.
- Highlight sequence detection.

to decouple the two words. They're not identical and never have been. DNA is a *thing*—a chemical that sticks to your fingers. Genes have a physical nature, too; in fact, they're made of long stretches of DNA. But in some ways genes are better viewed as conceptual, not material. A gene is really information—more like a story, with DNA as the language the story is written in. DNA and genes combine to form larger structures called chro-

What is DNA sequencing?

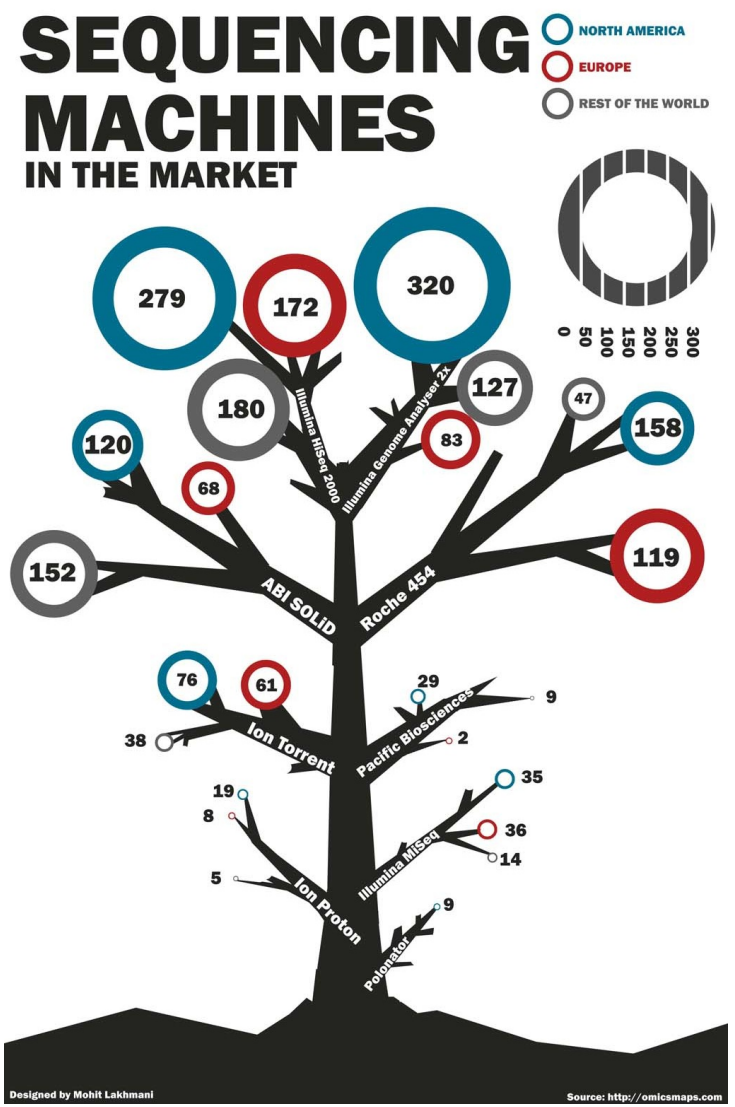
It is reading the letters of the book.
It is reading the exact nucleotide sequence of the genome.

What is DNA sequencing?

1. The identity of a nucleotide
2. The location of a nucleotide

Sequencing methods available now!

1. Maxam and Gilbert chemical degradation method (extinct).
2. Sanger sequencing (dideoxy or chain termination method).
3. Illumina sequencing.
4. SOLiD sequencing.
5. Pyrosequencing.
6. Ion Torrent method.
7. Single molecule sequencing.



Why DNA sequencing?

- DNA sequencing can be considered the ultimate characterization of gene(s) or fragment(s) of DNA.
- **DNA Sequencing is used for:**
  - Mapping genomes
  - Determining gene structure and thus function
  - Detecting polymorphism (single nucleotide polymorphism SNP)
  - Analyzing genetic variation
  - Predicting the possible product(s) of DNA fragments
  - Many purposes depending on the questions one is asking

DNA sequencing methods

When learning about DNA sequencing methods, focus on:

1. How the nucleotide identity is identified (e.g., synthesis)?
2. What is the method of detection and differentiation between the four nucleotides?

Maxam-Gilbert Sequencing

In 1977, Allan Maxam and Walter Gilbert developed chemical sequencing method

Proc. Natl. Acad. Sci. USA
Vol. 74, No. 2, pp. 560–564, February 1977
Biochemistry

A new method for sequencing DNA

(DNA chemistry/dimethyl sulfate cleavage/hydrazine/piperidine)

ALLAN M. MAXAM AND WALTER GILBERT

Department of Biochemistry and Molecular Biology, Harvard University, Cambridge, Massachusetts 02138

Contributed by Walter Gilbert, December 9, 1976

Maxam-Gilbert Sequencing

1. Base modification.
2. Removal of modified base from its sugar.
3. Breaking the phosphodiester bond.
4. Analyzing the fragments using gel electrophoresis.

Maxam-Gilbert Sequencing

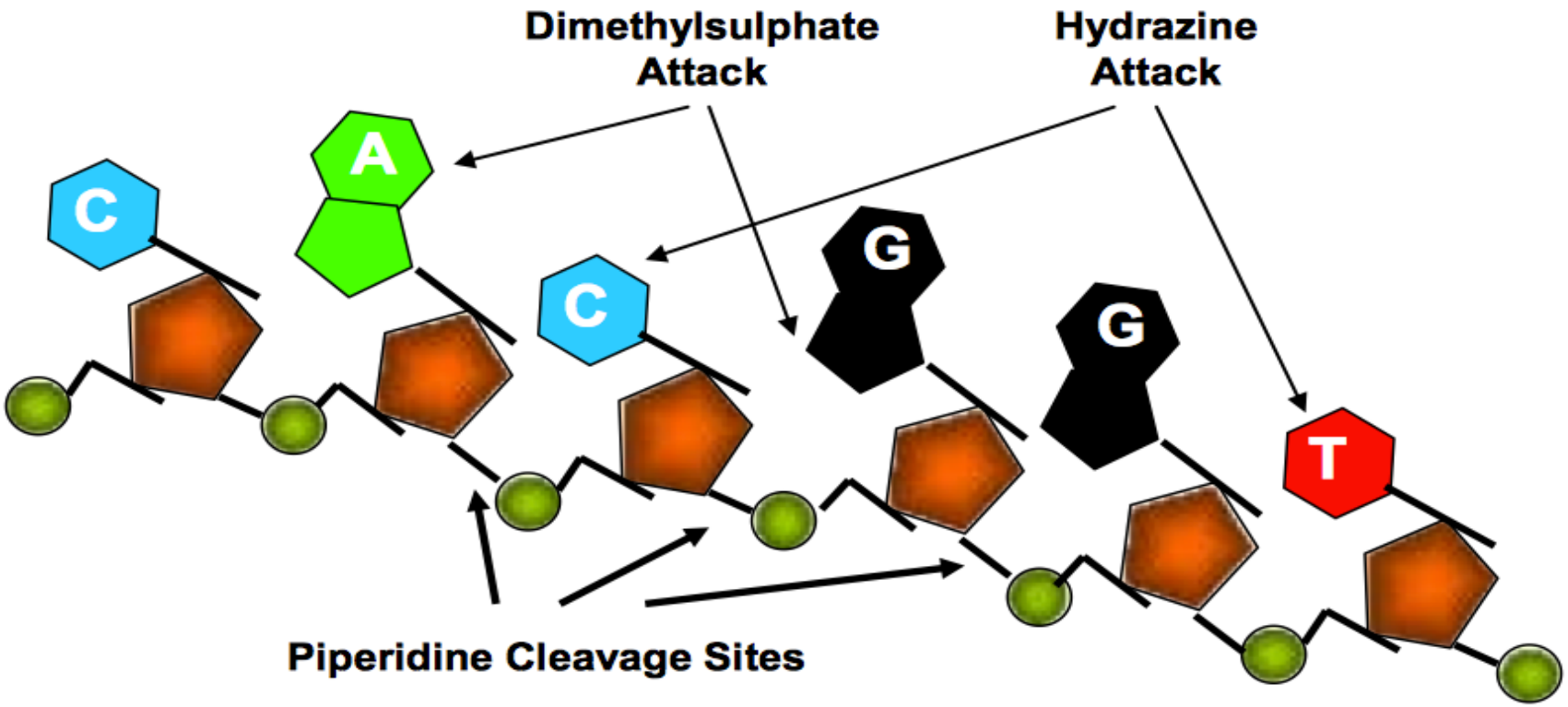


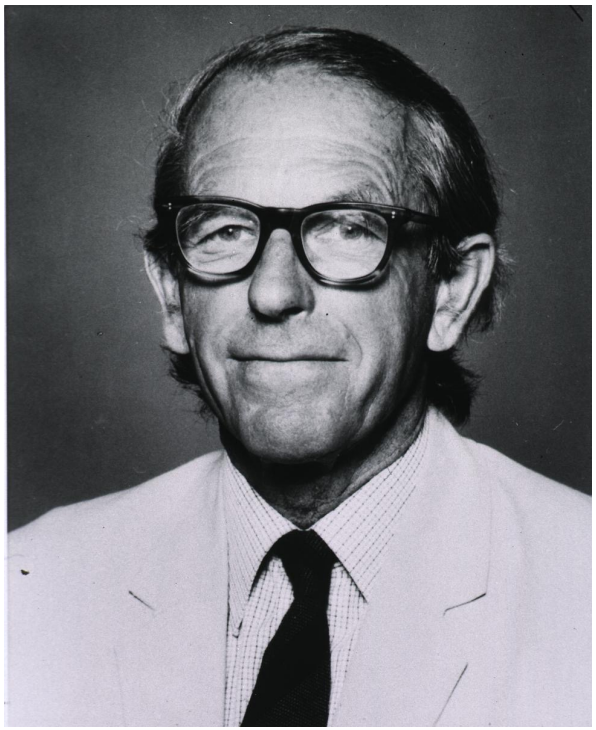
Figure 1. Chemical targets in the Maxam-Gilbert DNA sequencing strategy. Dimethylsulphate or hydrazine will attack the purine or pyrimidine rings respectively and piperidine will cleave the phosphate bond at the 3' carbon.

Maxam-Gilbert Sequencing

Any idea why no (T) or (A) lanes?

How do we read the gel?

Sanger sequencing



1. Fredrick Sanger has developed a sequencing method and received a Noble prize for it.

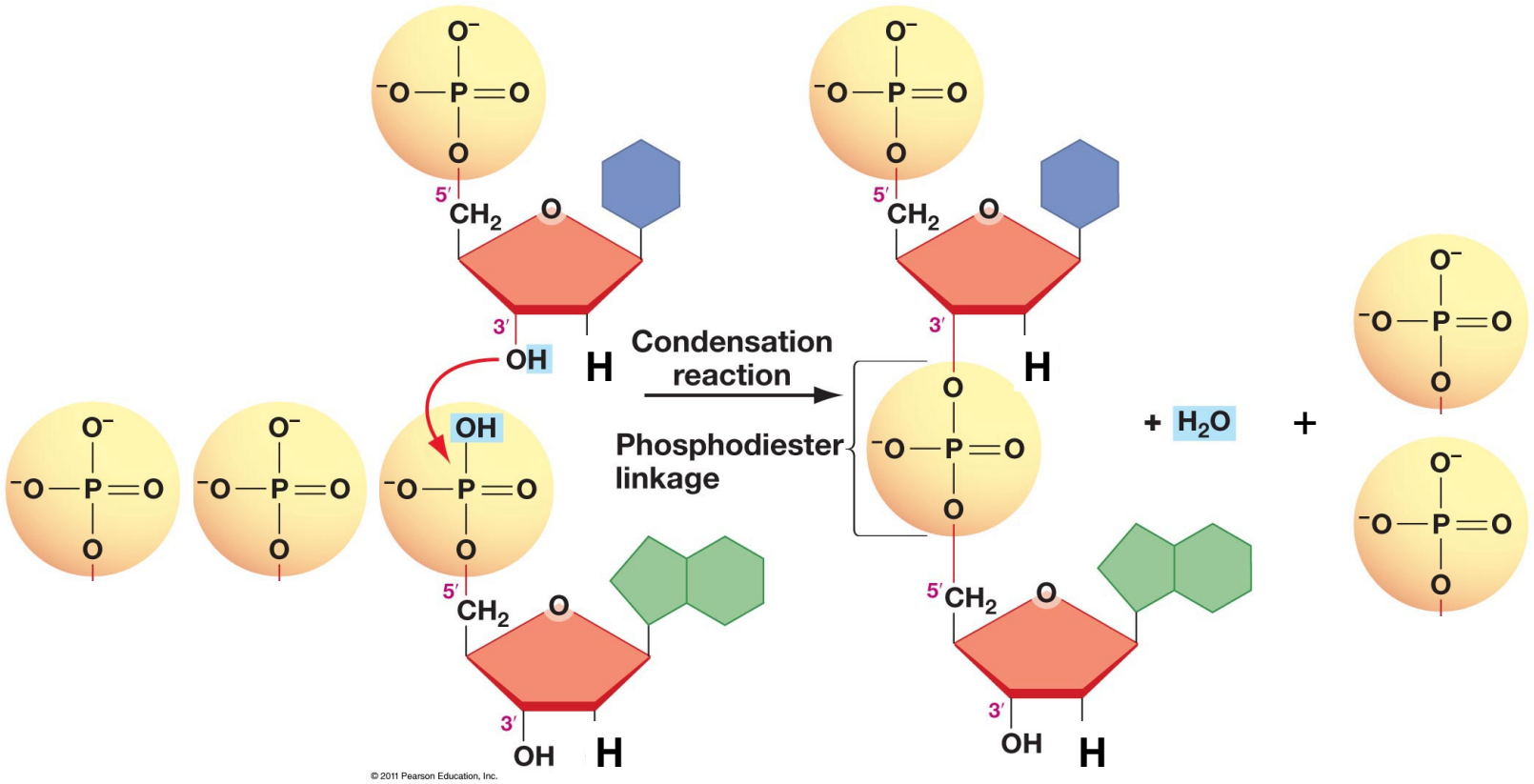
2. Sanger sequencing method is also called **Chain Termination Method** and **Dideoxy sequencing method.**

Sanger sequencing

- **Employs:**
  - specific primers
  - dNTPs
  - ddNTPs
  - DNA polymerase
  - DNA template

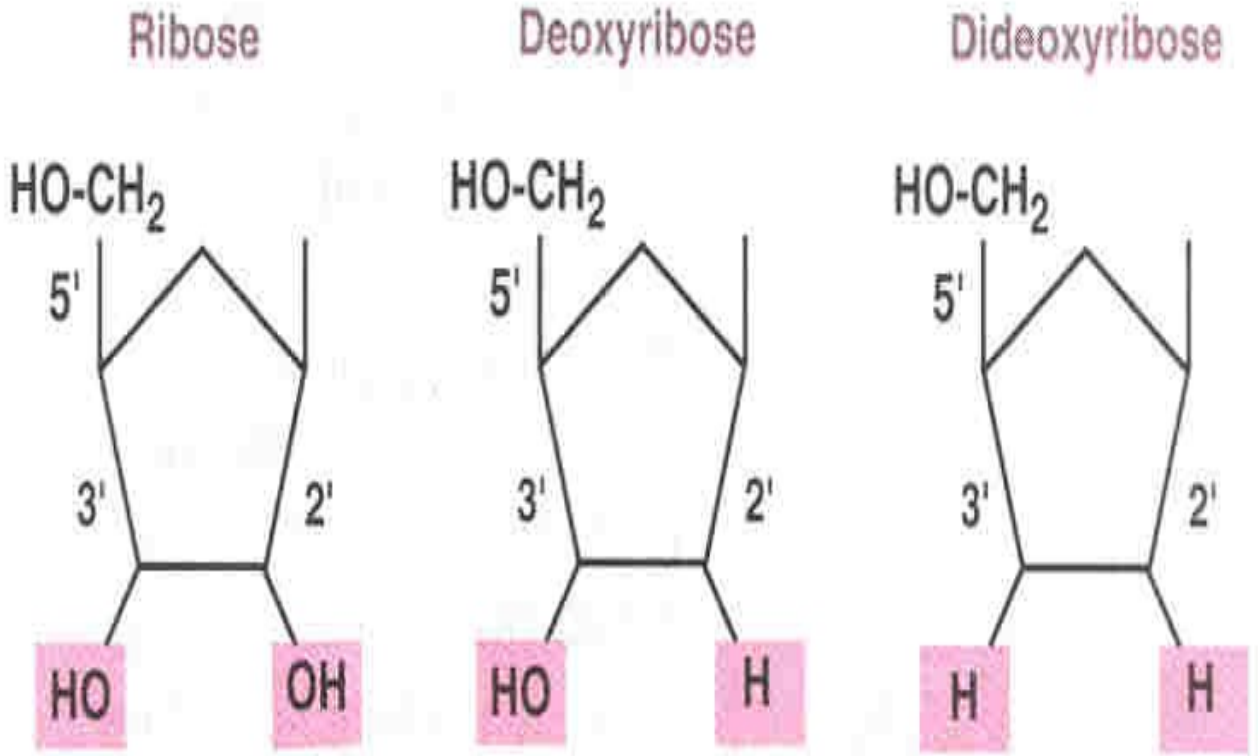
DNA synthesis

DNA synthesis requires the availability of a 3’-OH and energy



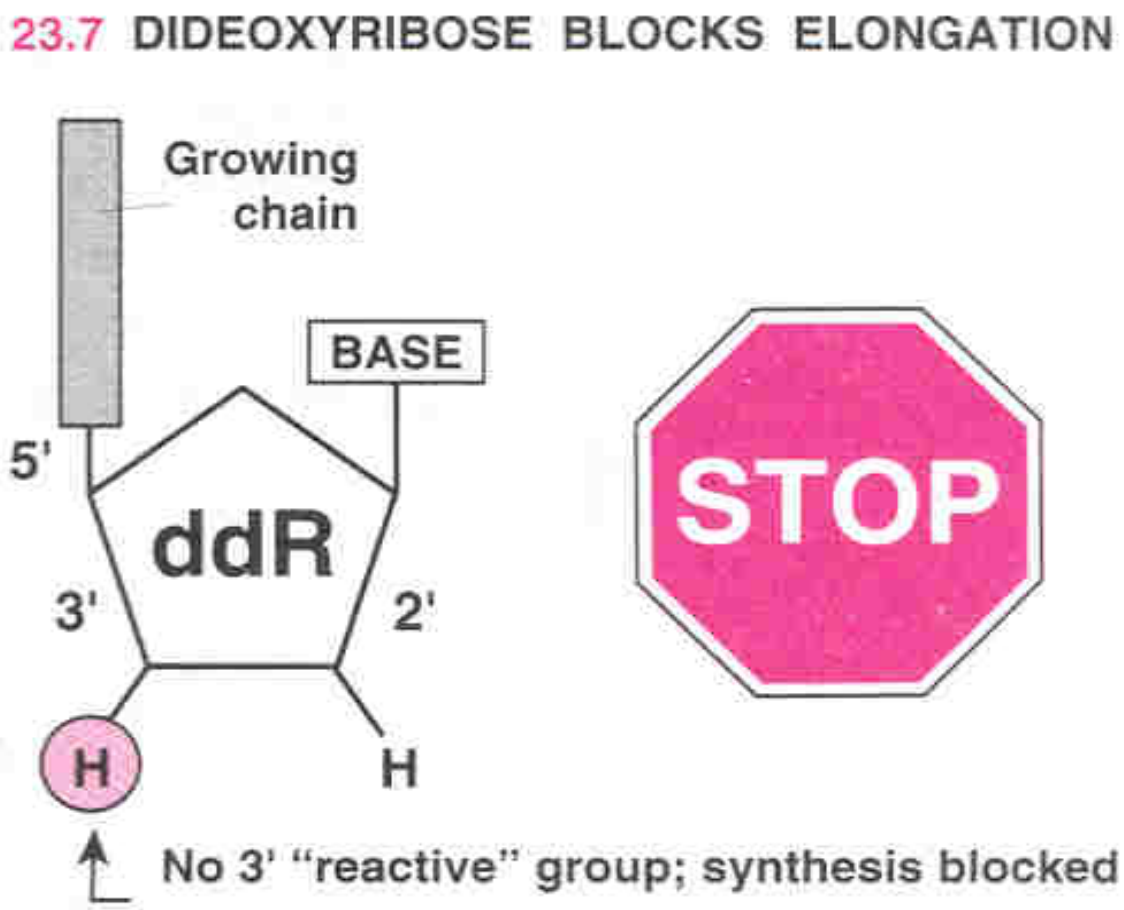
DNA synthesis

Difference in OH location in sugar and consequences



DNA synthesis

The absence of OH group on the 3’ carbon of the sugar blocks further addition of nucleotides



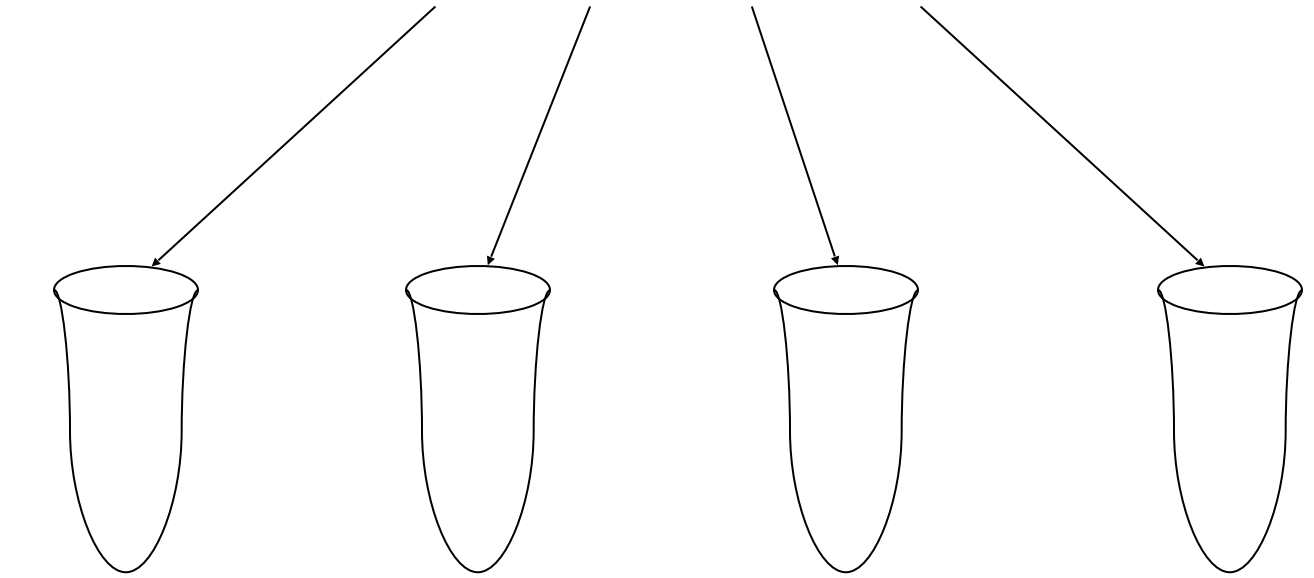
Sanger sequencing procedure

DNA Template

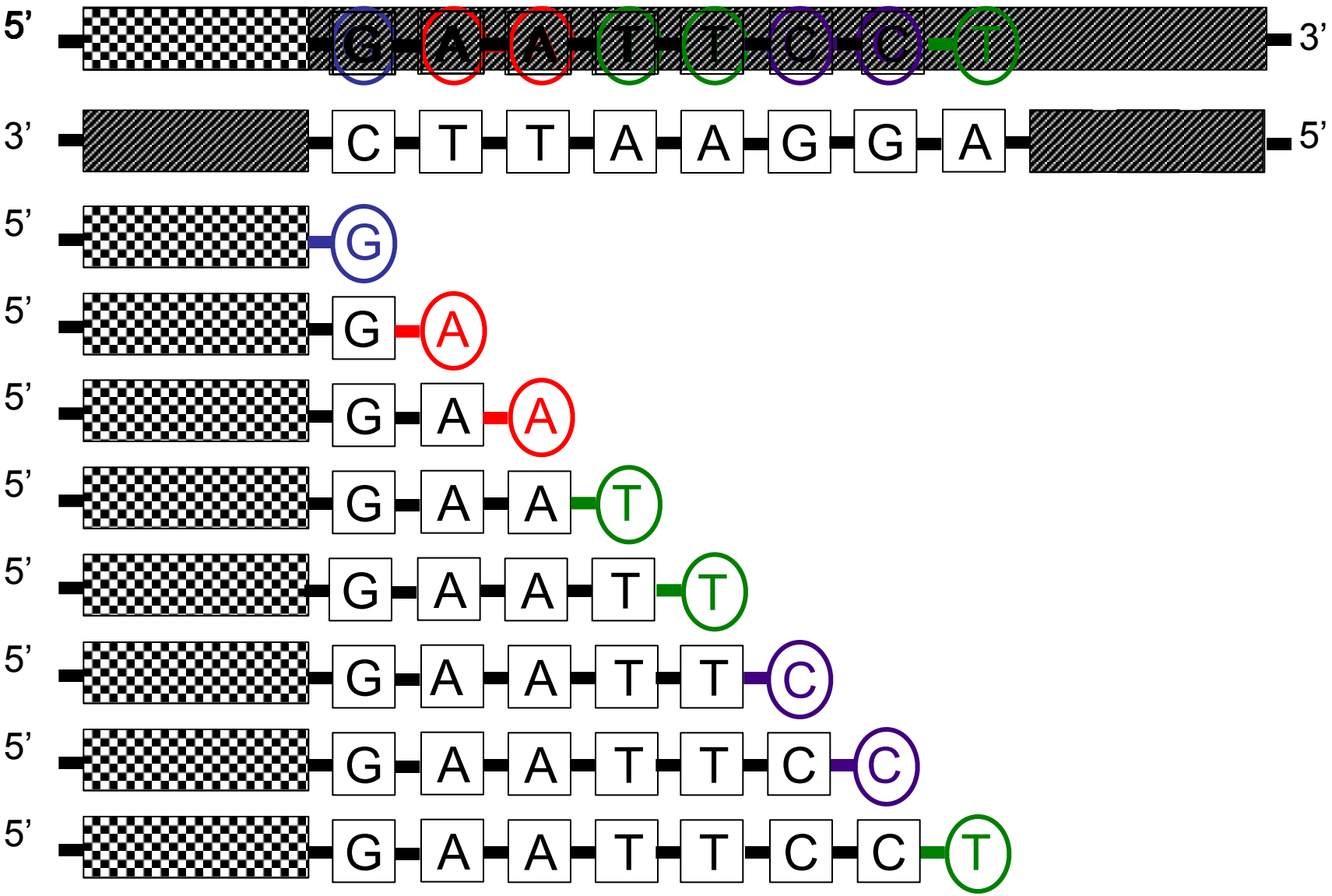
Polymerase

Excess dNTPs

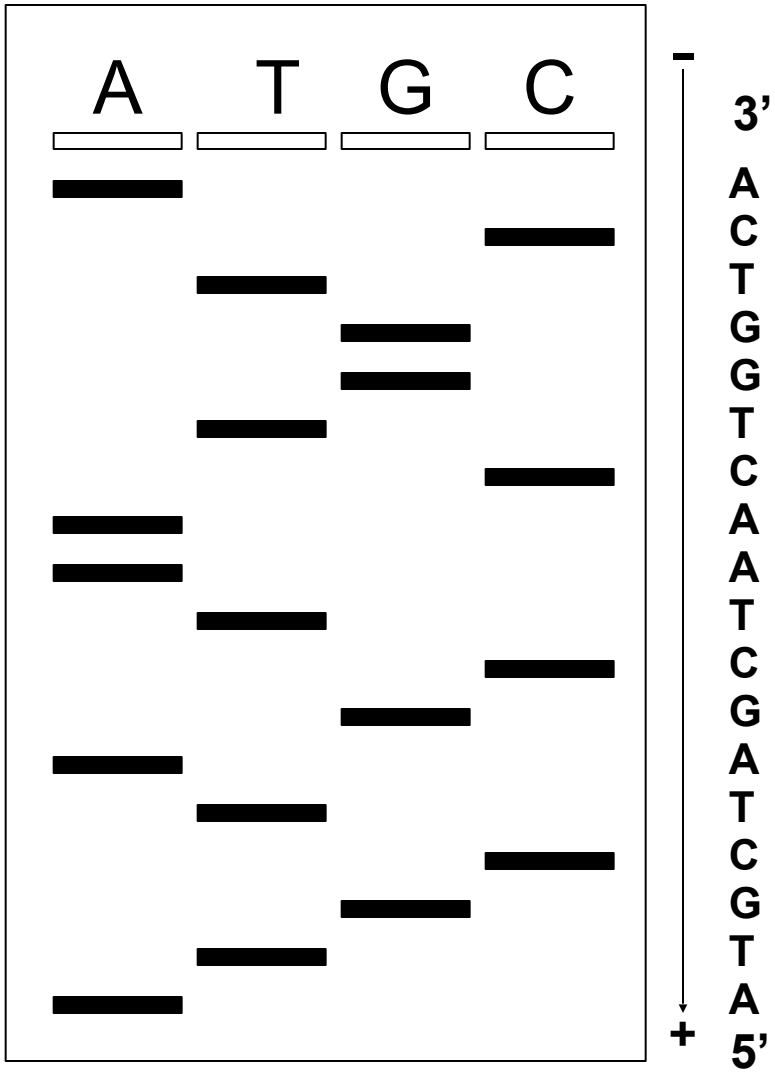
Primer



Sanger sequencing procedure

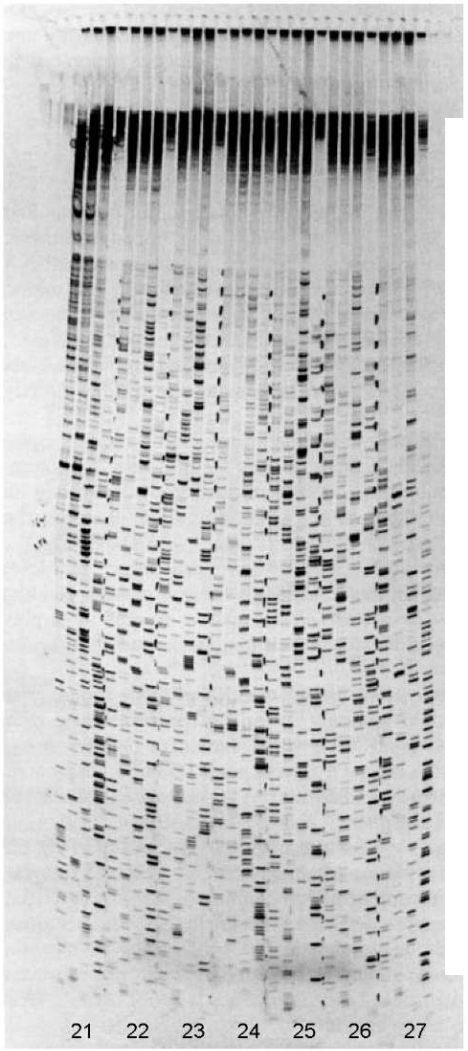


Sanger sequencing procedure



- Analysis using high resolution polyacrylamide gel electrophoresis.
- Fragments are detected using radioactive markers and autoradiography.

Sanger sequencing - Gel



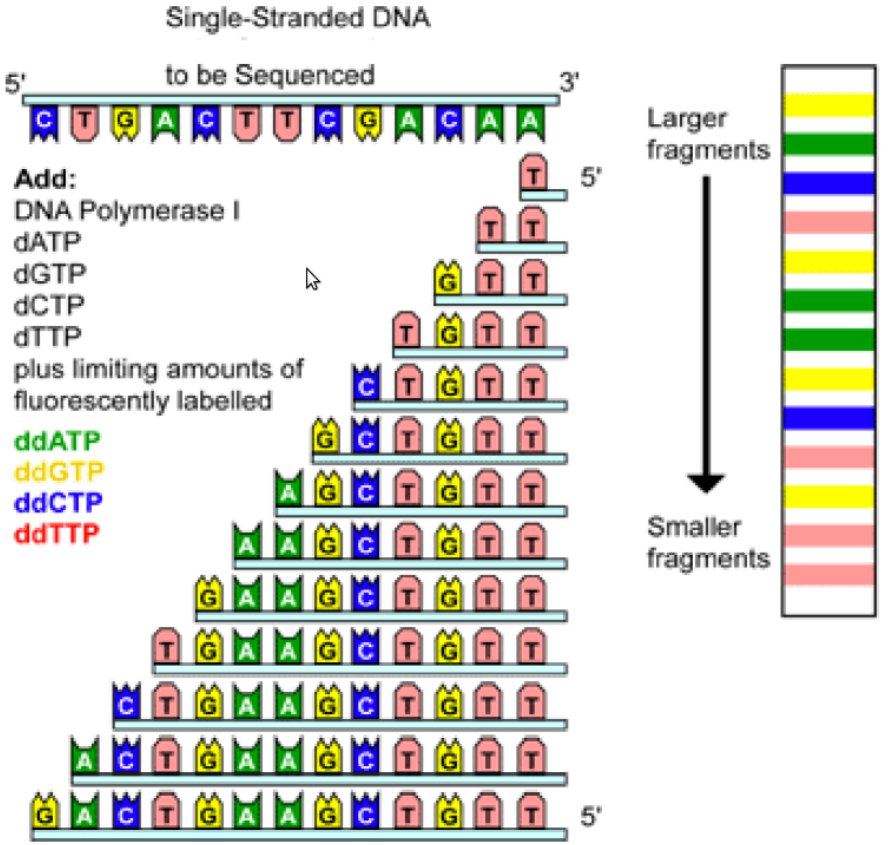
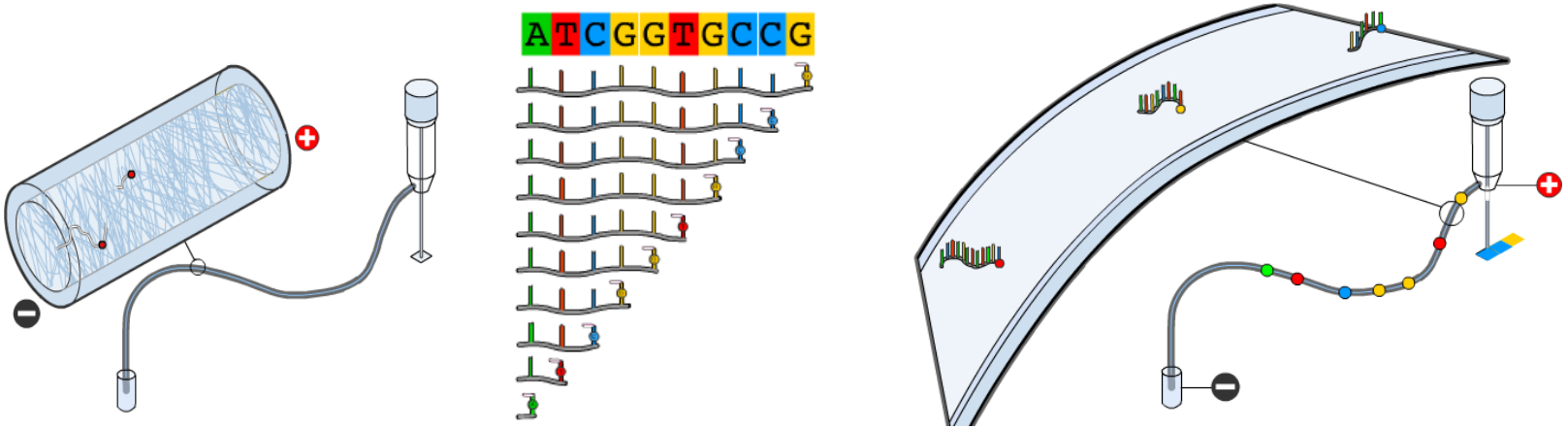


FIGURE 2.8. An autoradiogram (X-ray film) of a DNA sequencing gel. Each sequence requires Four lanes, one for each base.

Sanger sequencing - Automated

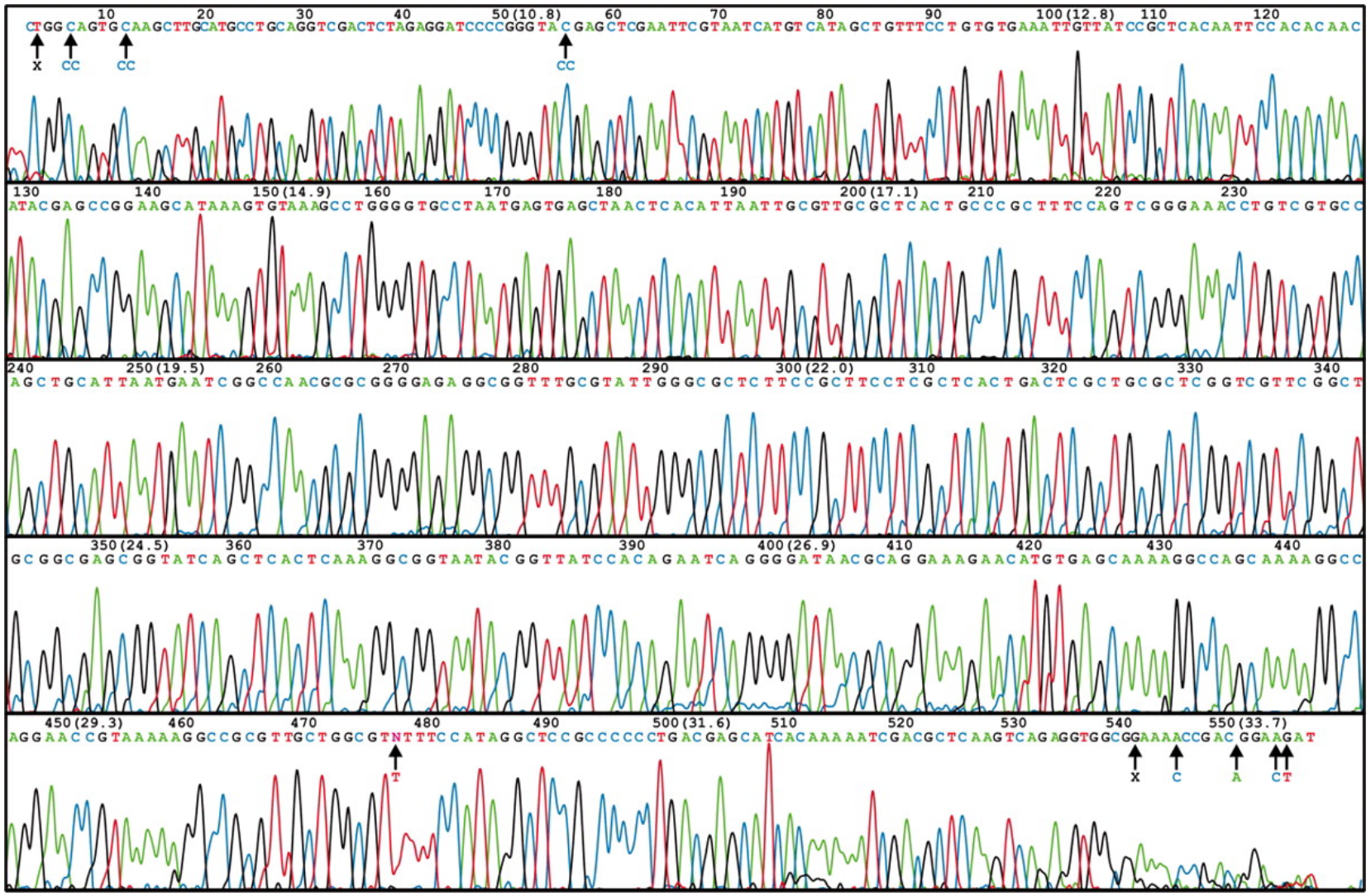
• Each dideoxy nucleotide is attached to a florescent marker.

• At the end of each cycle, a laser beam can detect the florescent marker and thus record the position of the nucleotide.



Sanger sequencing - Automated

Chromatogram - Automated



The so called NGS

Current sequencing technologies

Next-generation sequencing must die!

“NGS technologies have already been around for almost 20 years. It doesn't strike me as particularly helpful to keep on labeling all of these different technologies with the same name.” Keith Bradnam.

http://www.acgt.me/blog/2014/3/7/next-generation-sequencing-must-die

Current sequencing technologies

What should we call them?

- Stop using NGS to describe sequencing technology.
- Refer to the technology by its method (e.g. sequencing-by-synthesis).
- Refer to the technology by the company that provides it (e.g. Oxford Nanopore).
- Use the term “current sequencing technologies” if people can know what current is.

What is “current” to you today?

http://www.acgt.me/blog/2014/3/7/next-generation-sequencing-must-die

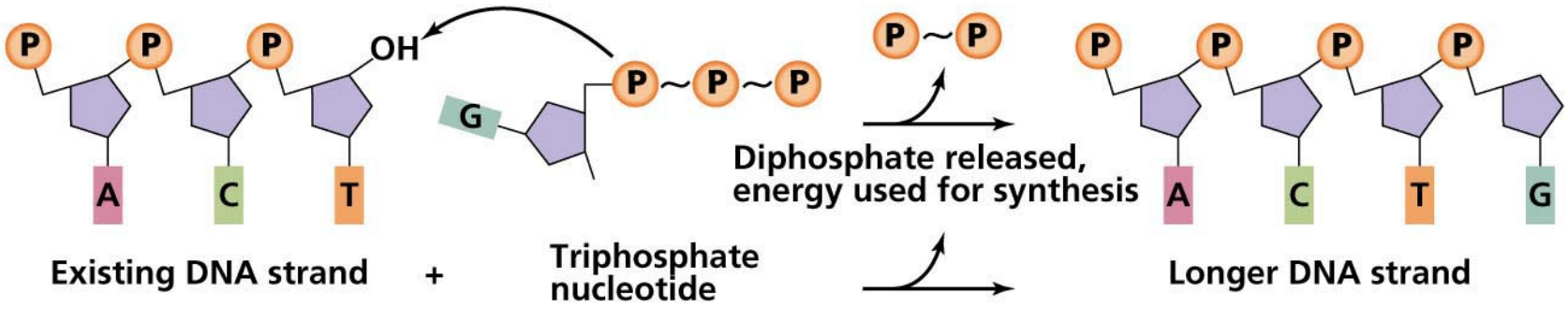
Pyrosequencing

- Method developed by Pål Nyrén and Mustafa Ronaghi in 1996.
- It is a DNA sequencing method by synthesis.
- The method detects incorporation of nucleotides by the pyrophosphate by product.

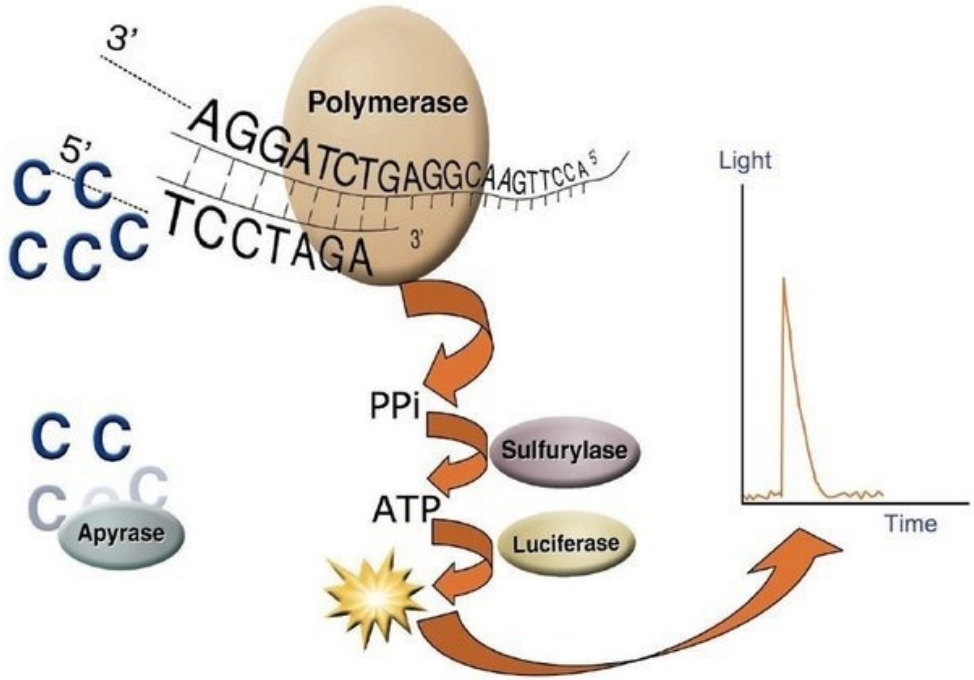
Pyrosequencing Principle

What is pyrophosphate?

Pyrosequencing Principle



(b)

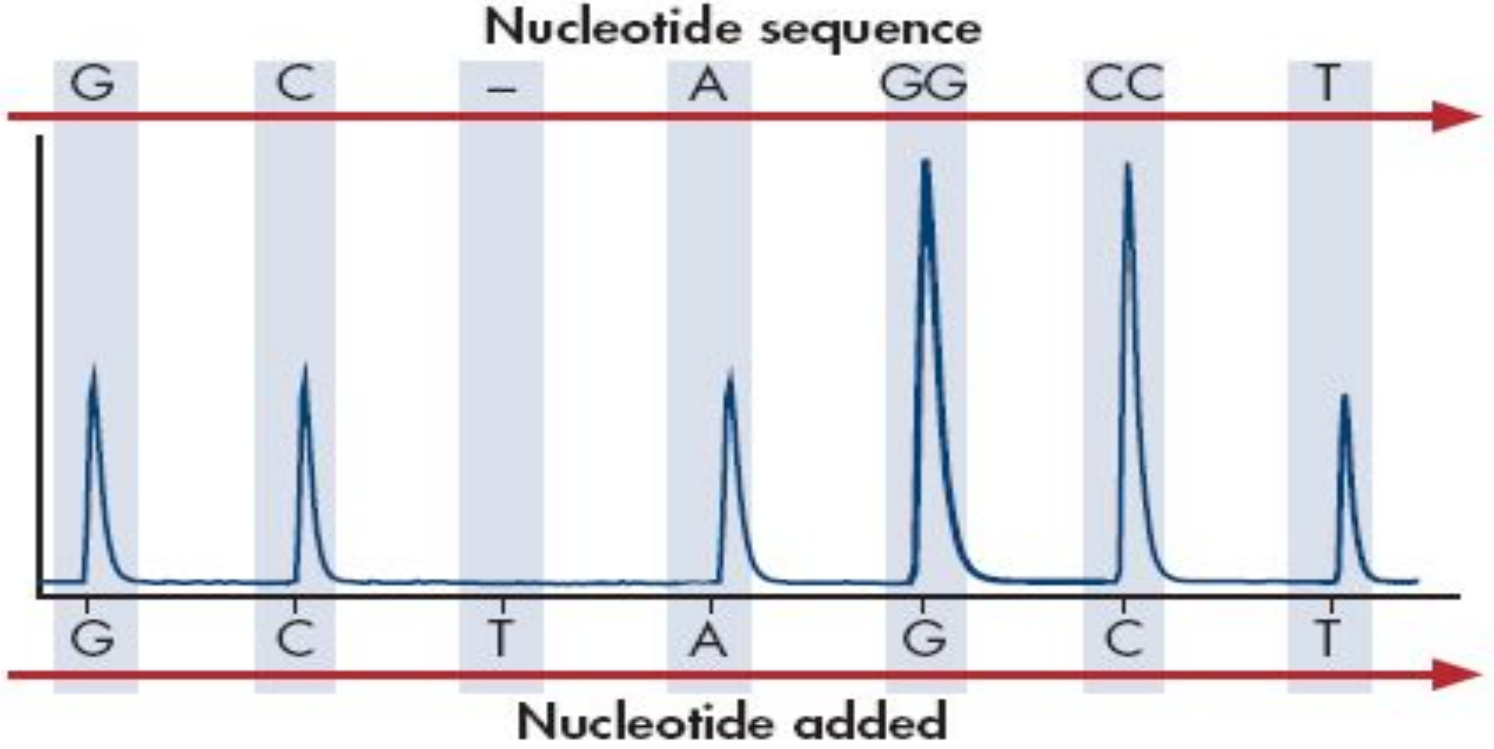


The enzyme cascade of the Pyrosequencing technology.

Pyrosequencing Principle

What does the resulting sequence look like?

Pyrogram

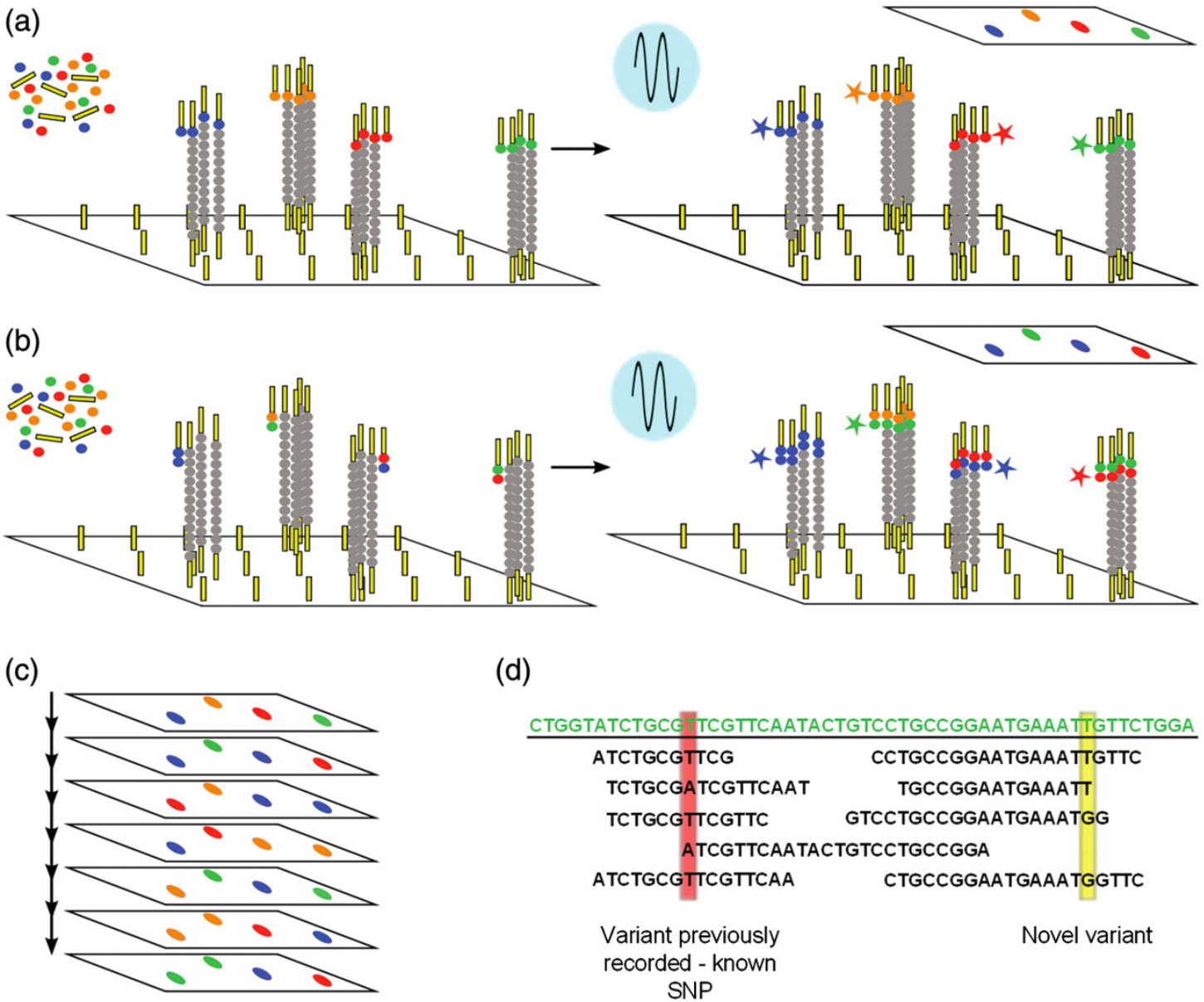


Illumina Sequencing technology

- It is a DNA sequencing method by synthesis.
- The method detects dNTP incorporation using fluorescence labeled nucleotides.
- Incorporation of label dNTPs terminates synthesis.
- After detection(imaging), the label is removed to allow incorporation of next nucleotide.

https://www.illumina.com/documents/products/techspotlights/techspotlight_sequencing.pdf

Illumina Sequencing technology

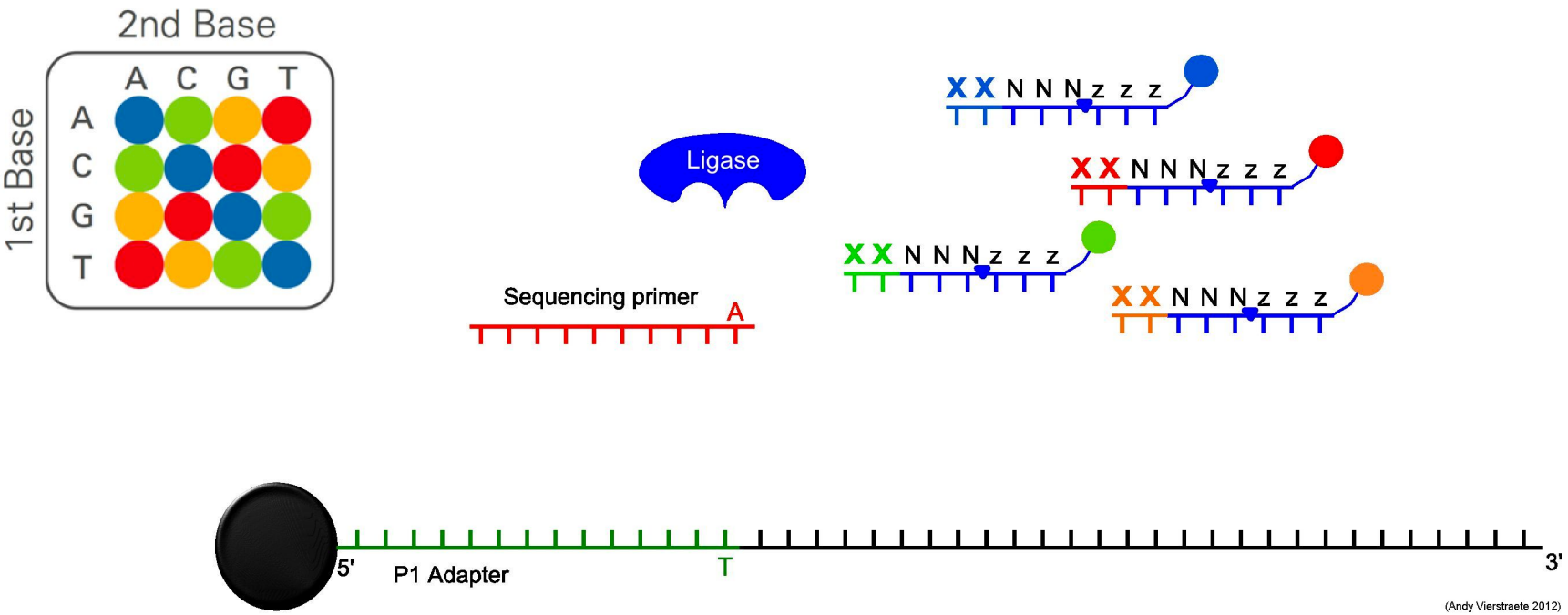


ABi SOLiD Sequencing

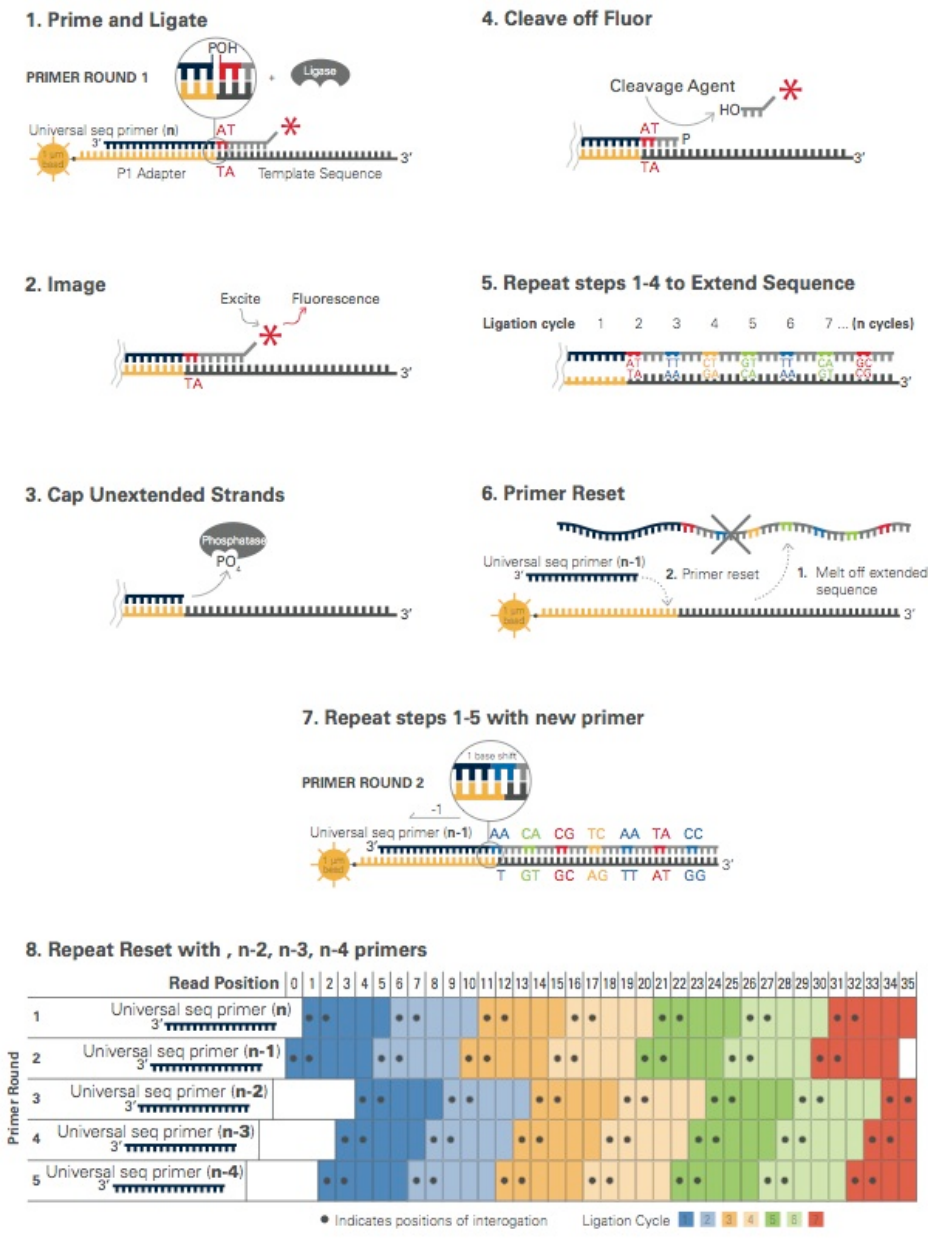
- SOLiD stands for (Sequencing by Oligonucleotide Ligation and Detection).
- Sequencing by ligation.
- Ligation of four fluorescent labeled di-base probes.
- Different universal primers (different lengths) are used.
- The overlap of all ligation rounds will determine the DNA sequence.

http://www.appliedbiosystems.com/absite/us/en/home/applications-technologies/solid-next-generation-sequencing/next-generation-systems/solid-sequencing-chemistry.html

ABi SOLiD Sequencing

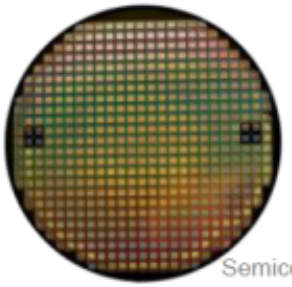


ABi SOLiD Sequencing

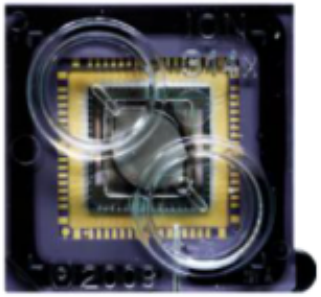


Ion Torrent sequencing

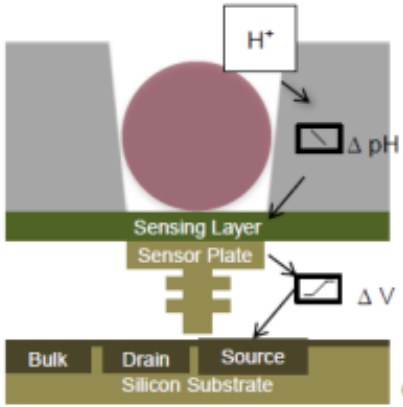
- A semi-conductor sequencing technology.
- Small chips contain millions of wells/sensors.
- Does not use fluorescent light or camera imaging.



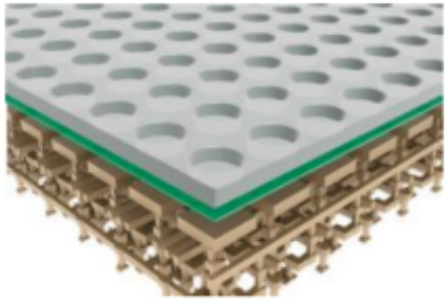
Wafer
Semiconductor Manufacturing



Chip
Semiconductor Packaging



Single Sensor
Chemical to Digital Sequence

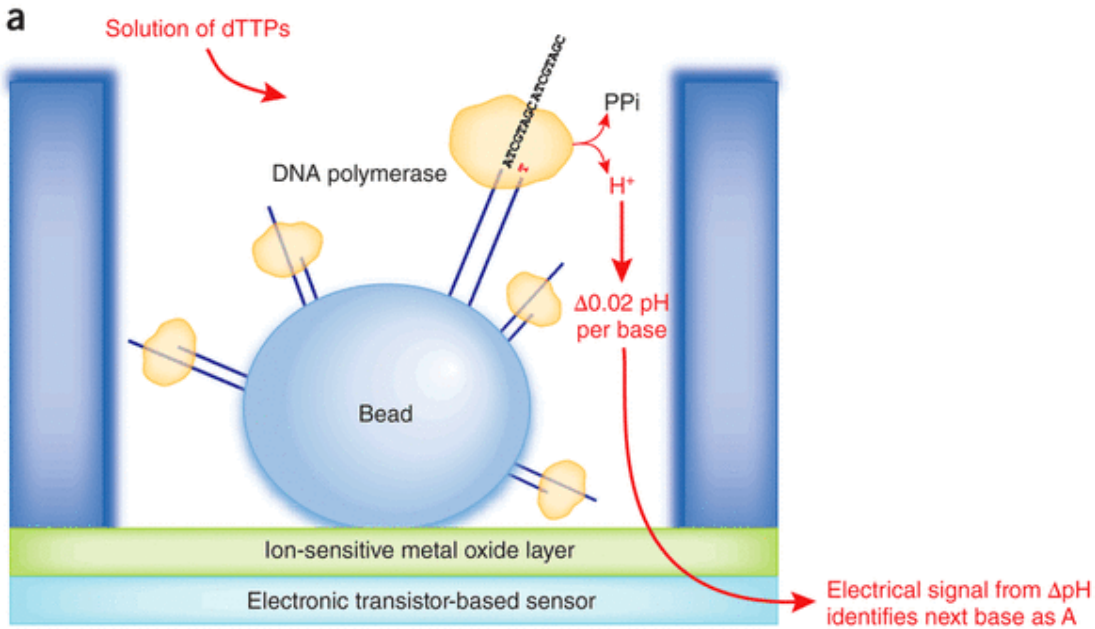


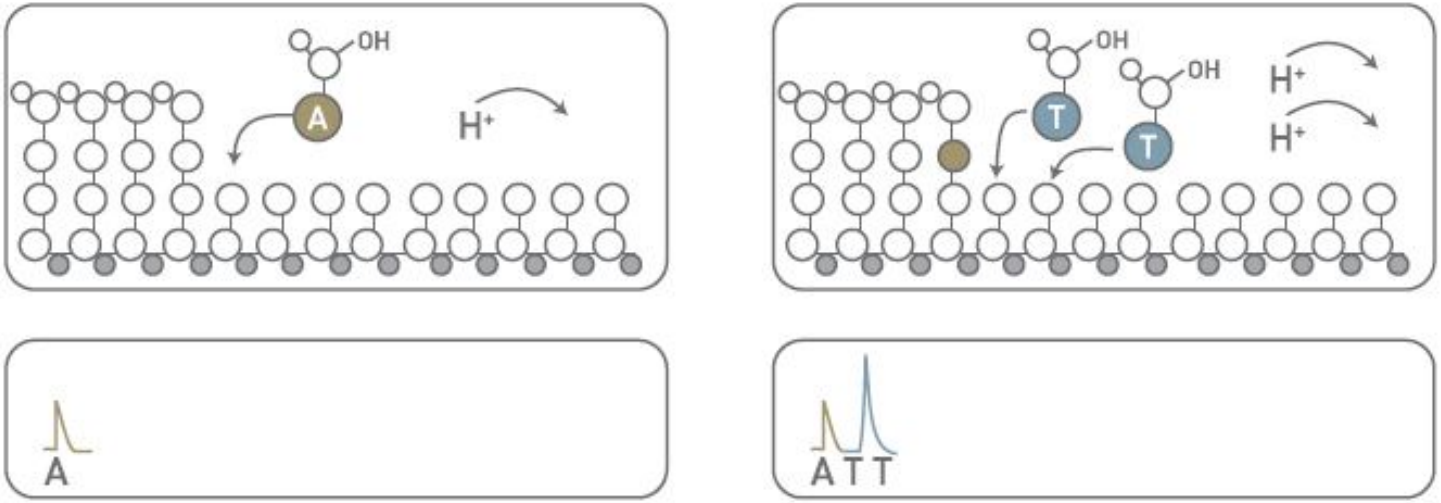
Millions of Sensors
Semiconductor Design

Principle

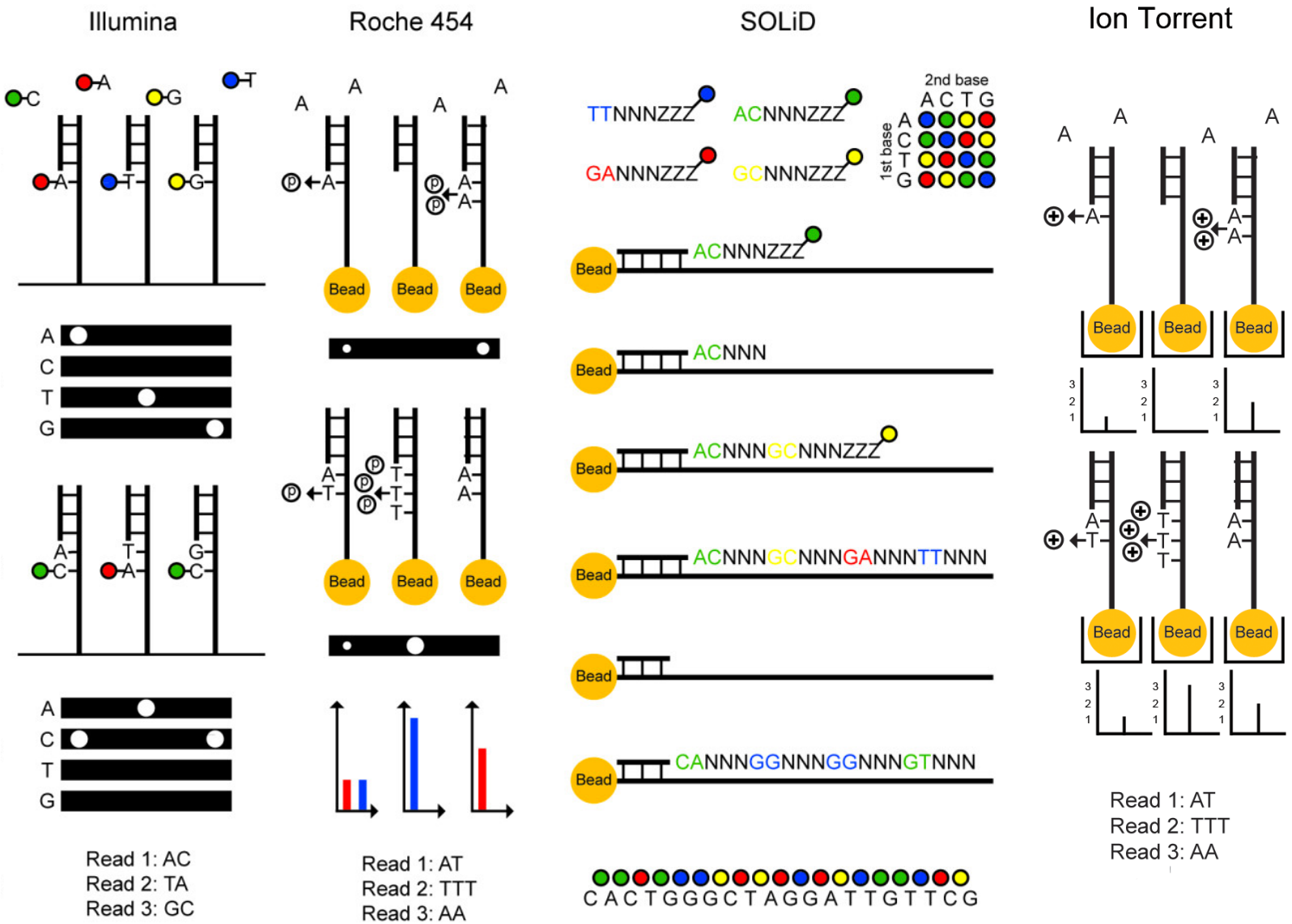
- Incorporation of correct nucleotide causes changes in pH.
- Why?
- Ion sensor that set under the well can detects\measures the changes of the pH.
- The chemical signals and converts it to voltage.
- This process is repeated every 15 seconds with different nucleotides washing over the chip.

Principle


a
Solution of dTTPs
DNA polymerase
PPi
H+
Δ0.02 pH per base
Bead
Ion-sensitive metal oxide layer
Electronic transistor-based sensor
Electrical signal from ΔpH identifies next base as A


OH
A
H+
OH
OH
T
T
H+
H+
A
A T T

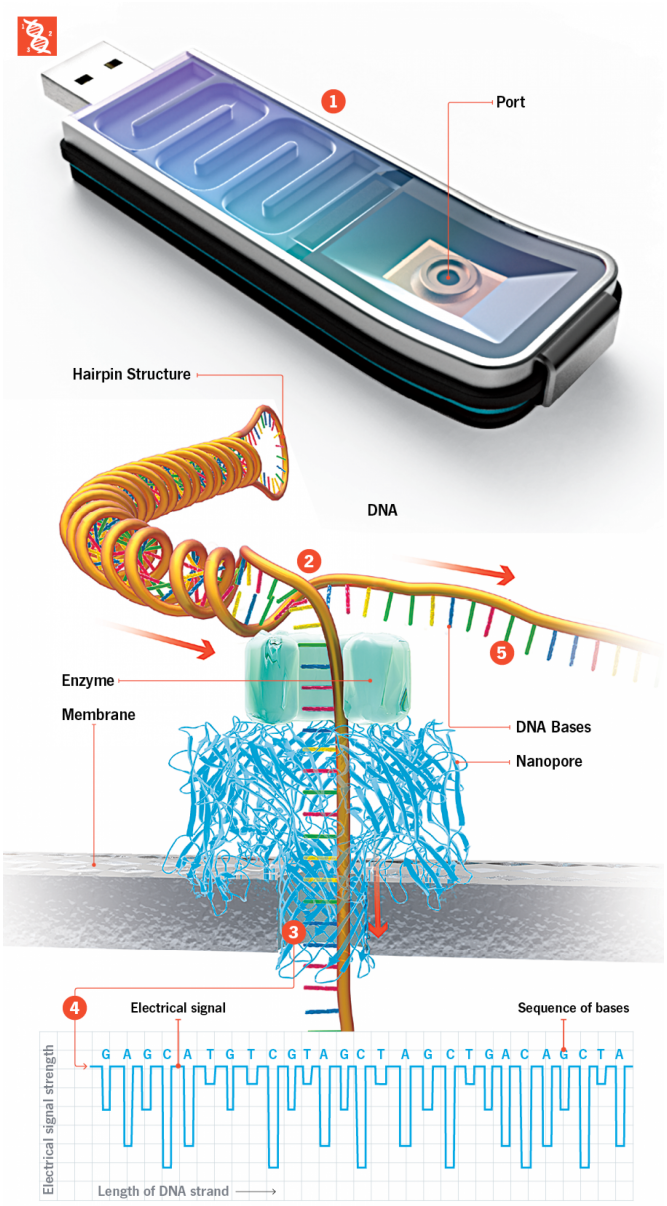
Nice comparison



Nanopore sequencing

- No Synthesis or ligation.
- Realtime sequencing.
- Uses nanopores.
- Changes in electrical current passing through the nanopore is the method of nucleotide detection and identification.
- Each nucleotide has a specific electrical current disruption signature.

Principle



1. Double stranded DNA attaches to a helicase.

2. Single strand DNA passes through the nanopore and changes in flow of electric current is detected.

3. Each base has a different electrical signal.

Fast - Realtime - Accurate - Small device - Cheap
The future?

Minion

MinION Mkl: portable, real time biological analyses

MinION

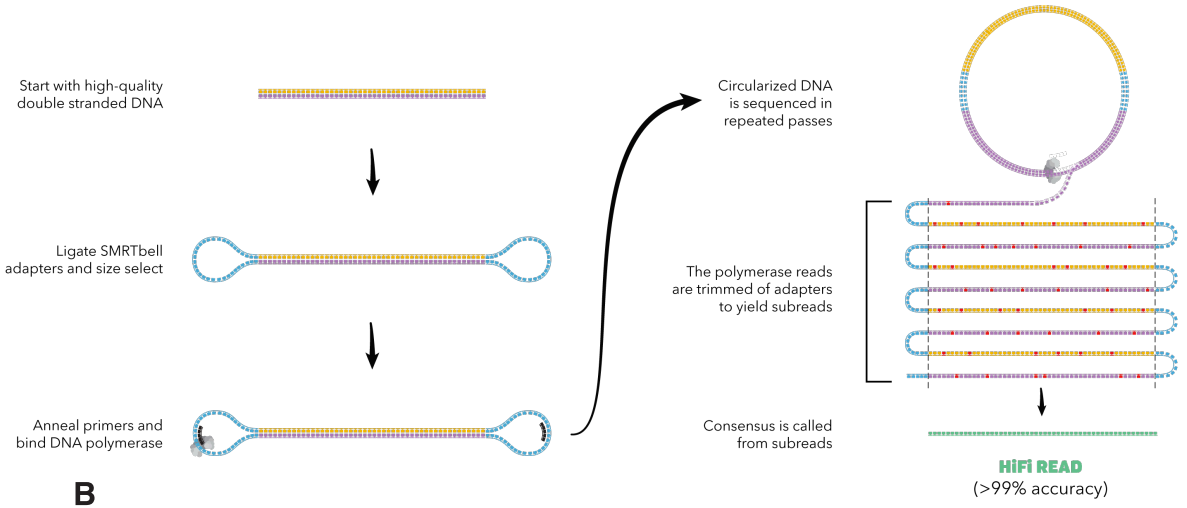
PacBio sequencing

- Real time sequencing by synthesis.
- No pause between steps.
- Generate longer read lengths.
- Higher error rate!
- A template (SMRTbell) is created by ligating hairpin adaptors to both ends of (dsDNA).
- Four different fluorescent-labeled nucleotides
- Distinguish between fluorescent signals

PacBio sequencing

Real-Time DNA Sequencing from Single Polymerase Molecules

SCIENCE VOL 323 2 JANUARY 2009



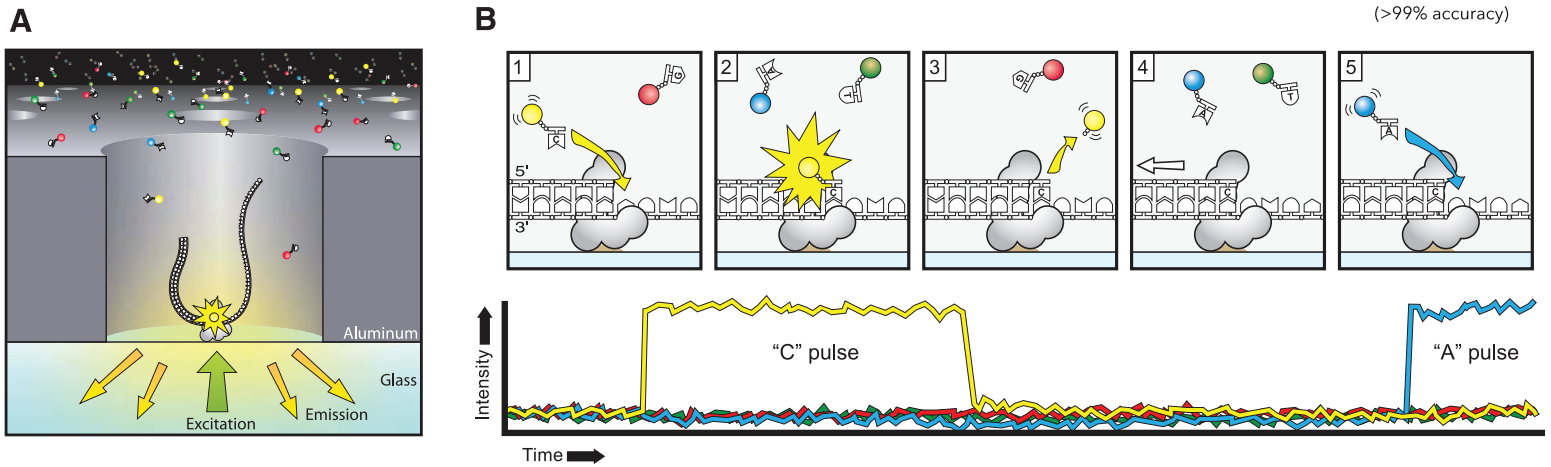


Fig. 1. Principle of single-molecule, real-time DNA sequencing. (**A**) Experimental geometry. A single molecule of DNA template-bound Φ29 DNA polymerase is immobilized at the bottom of a ZMW, which is illuminated from below by laser light. The ZMW nanostructure provides excitation confinement in the zeptoliter (10^{-21} liter) regime, enabling detection of individual phospholinked nucleotide substrates against the bulk solution background as they are incorporated into the DNA strand by the polymerase. (**B**) Schematic event sequence of the phospholinked dNTP incorporation cycle, with a corresponding expected time trace of detected fluorescence intensity from the ZMW. (1) A phospholinked nucleotide forms a cognate association with the template in the polymerase active site, (2) causing an elevation of the fluorescence output on the corresponding color channel. (3) Phosphodiester bond formation liberates the dye-linker-pyrophosphate product, which diffuses out of the ZMW, thus ending the fluorescence pulse. (4) The polymerase translocates to the next position, and (5) the next cognate nucleotide binds the active site beginning the subsequent pulse.

Which one to choose?

I suggest you look at this website for nice comparisons between sequencing technologies

http://www.molecularecologist.com/next-gen-fieldguide-2014/

Which one to choose?

Instrument	mplificatio	Run time	Millions of Reads/run	Bases / read	Reagent Cost/run	Reagent Cost/Gb	Reagent Cost/Mread	bp/run	Gbp/run	cost/Gb
Applied Biosystems 3730 (capillary)	PCR, cloning	2 hrs.	0.000096	650	$144	$2,307,692.	$1,500,000.	62,400	0.000	$2,307,692.3
454 GS Jr. Titanium	emPCR	10 hrs.	0.1	400	$977	$19,540.00	$9,770.00	50,000,000	0.050	$19,540.00
454 FLX Titanium	emPCR	10 hrs.	1	400	$6,200	$15,500.00	$6,200.00	400,000,000	0.400	$15,500.00
454 FLX+	emPCR	20 hrs.	1	650	$6,200	$9,538.46	$6,200.00	650,000,000	0.650	$9,538.46
Illumina GA IIx - v5 SE	bridgePCR	2 days	640	36	$4,842	$210.16	$7.57	23,040,000,000	23.040	$210.16
Illumina GA IIx - v5 PE	bridgePCR	14 days	640	288	$17,978	$97.54	$28.09	184,320,000,000	184.320	$97.54
Illumina MiSeq v2 Nano	bridgePCR	17 hrs.	1	300	$530	$1,766.67	$530.00	300,000,000	0.300	$1,766.67
Illumina MiSeq v2 Nano	bridgePCR	28 hrs.	1	500	$639	$1,278.00	$639.00	500,000,000	0.500	$1,278.00
Illumina MiSeq v2 Micro	bridgePCR	19 hrs.	4	300	$798	$665.00	$199.50	1,200,000,000	1.200	$665.00
Illumina MiSeq v2	bridgePCR	5 hrs.	15	50	$747	$996.00	$49.80	750,000,000	0.750	$996.00
Illumina MiSeq v2	bridgePCR	24 hrs.	15	300	$958	$212.89	$63.87	4,500,000,000	4.500	$212.89
Illumina MiSeq v2	bridgePCR	39 hrs.	15	500	$1,066	$142.13	$71.07	7,500,000,000	7.500	$142.13
Illumina MiSeq v3	bridgePCR	20 hrs.	22	150	$824	$249.70	$37.45	3,300,000,000	3.300	$249.70
Illumina MiSeq v3	bridgePCR	55 hrs.	22	600	$1,442	$109.24	$65.55	13,200,000,000	13.200	$109.24
Illumina NextSeq 500	BridgePCR	15 hrs.	130	150	$975	$50.00	$7.50	19,500,000,000	19.500	$50.00
Illumina NextSeq 500	BridgePCR	26 hrs.	130	300	$1,560	$40.00	$12.00	39,000,000,000	39.000	$40.00
Illumina NextSeq 500	BridgePCR	11 hrs.	400	75	$1,300	$43.33	$3.25	30,000,000,000	30.000	$43.33
Illumina NextSeq 500	BridgePCR	18 hrs.	400	150	$2,500	$41.67	$6.25	60,000,000,000	60.000	$41.67
Illumina NextSeq 500	BridgePCR	30 hrs.	400	300	$4,000	$33.33	$10.00	120,000,000,000	120.000	$33.33
Illumina HiSeq 2500 - rapid run	BridgePCR	10 hrs.	300	50	$1,350	$90.00	$4.50	15,000,000,000	15.000	$90.00
Illumina HiSeq 2500 - rapid run	BridgePCR	27 hrs.	300	200	$3,126	$52.10	$10.42	60,000,000,000	60.000	$52.10
Illumina HiSeq 2500 - rapid run	BridgePCR	40 hrs.	300	300	$4,126	$45.84	$13.75	90,000,000,000	90.000	$45.84
Illumina HiSeq 2500 - high output v3	BridgePCR	2 days	1500	50	$5,866	$78.21	$3.91	75,000,000,000	75.000	$78.21
Illumina HiSeq 2500 - high output v3	BridgePCR	11 days	1500	200	$13,580	$45.27	$9.05	300,000,000,000	300.000	$45.27
Illumina HiSeq 2500 - high output v4	BridgePCR	40 hrs.	2000	50	$5,866	$58.66	$2.93	100,000,000,000	100.000	$58.66
Illumina HiSeq 2500 - high output v4	BridgePCR	6 days	2000	250	$14,950	$29.90	$7.48	500,000,000,000	500.000	$29.90
Illumina HiSeq X (2 flow cells)	BridgePCR	3 days	6000	300	$12,750	$7.08	$2.13	1,800,000,000,00	1,800.000	$7.08
Ion Torrent – PGM 314 chip	emPCR	2.3 hrs.	0.475	200	$349	$3,673.68	$734.74	95,000,000	0.095	$3,673.68
Ion Torrent – PGM 314 chip	emPCR	3.7 hrs.	0.475	400	$474	$2,494.74	$997.89	190,000,000	0.190	$2,494.74
Ion Torrent – PGM 316 chip	emPCR	3 hrs.	2.5	200	$549	$1,098.00	$219.60	500,000,000	0.500	$1,098.00
Ion Torrent – PGM 316 chip	emPCR	4.9 hrs.	2.5	400	$674	$674.00	$269.60	1,000,000,000	1.000	$674.00
Ion Torrent – PGM 318 chip	emPCR	4.4 hrs.	4.75	200	$749	$788.42	$157.68	950,000,000	0.950	$788.42
Ion Torrent – PGM 318 chip	emPCR	7.3 hrs.	4.75	400	$874	$460.00	$184.00	1,900,000,000	1.900	$460.00
Ion Torrent - Proton I	emPCR	4 hrs.	70	175	$1,000	$81.63	$14.29	12,250,000,000	12.250	$81.63
Ion Torrent - Proton II (forecast)	emPCR	5 hrs.	280	175	$1,000	$20.41	$3.57	49,000,000,000	49.000	$20.41
Ion Torrent - Proton III (forecast)	emPCR	6 hrs.	500	175	$1,000	$11.43	$2.00	87,500,000,000	87.500	$11.43
Life Technologies SOLiD – 5500xl	emPCR	8 days	1410	110	$10,503	$67.72	$7.45	155,100,000,000	155.100	$67.72
Pacific Biosciences RS II	None - SMS	2 hrs.	0.03	3000	$100	$1,111.11	$3,333.33	90,000,000	0.090	$1,111.11
Oxford Nanopore MinION (forecast)	None - SMS	≤6 hrs.	0.1	9000	$900	$1,000.00	$9,000.00	900,000,000	0.900	$1,000.00
Oxford Nanopore GridION 2000 (forecast)	None - SMS	varies	4	10000	$1,500	$37.50	$375.00	40,000,000,000	40.000	$37.50
Oxford Nanopore GridION 8000 (forecast)	None - SMS	varies	10	10000	$1,000	$10.00	$100.00	100,000,000,000	100.000	$10.00

http://www.molecularecologist.com/next-gen-fieldguide-2014/

Expectations

- For each of the sequencing methods/ technology, you know:
  - The general principle.
  - Sequencing by?
  - Sequence detection by?

Acknowledgment

Molecular Biology class (Spring 2014):

Taiba Albarjes
Noor Alrumaidhi
Masumah Jassim
Elaf Jamal
Amna Al-rubayaa
Esraa Jassim Mayouf
Kareemah Al-Enezi
Mona Abdi
Shahad Aljandal

Disclaimer

Figures, photos, and graphs in my lectures are collected using google searches. I do not claim to have personally produced the material (except for some). I do cite only articles or books used. I thank all owners of the visual aid that I use and apologize for not citing each individual item. If anybody finds the inclusion of their material in my lectures a violation of their copy rights, please contact me via email.

hhalhaddad@gmail.com