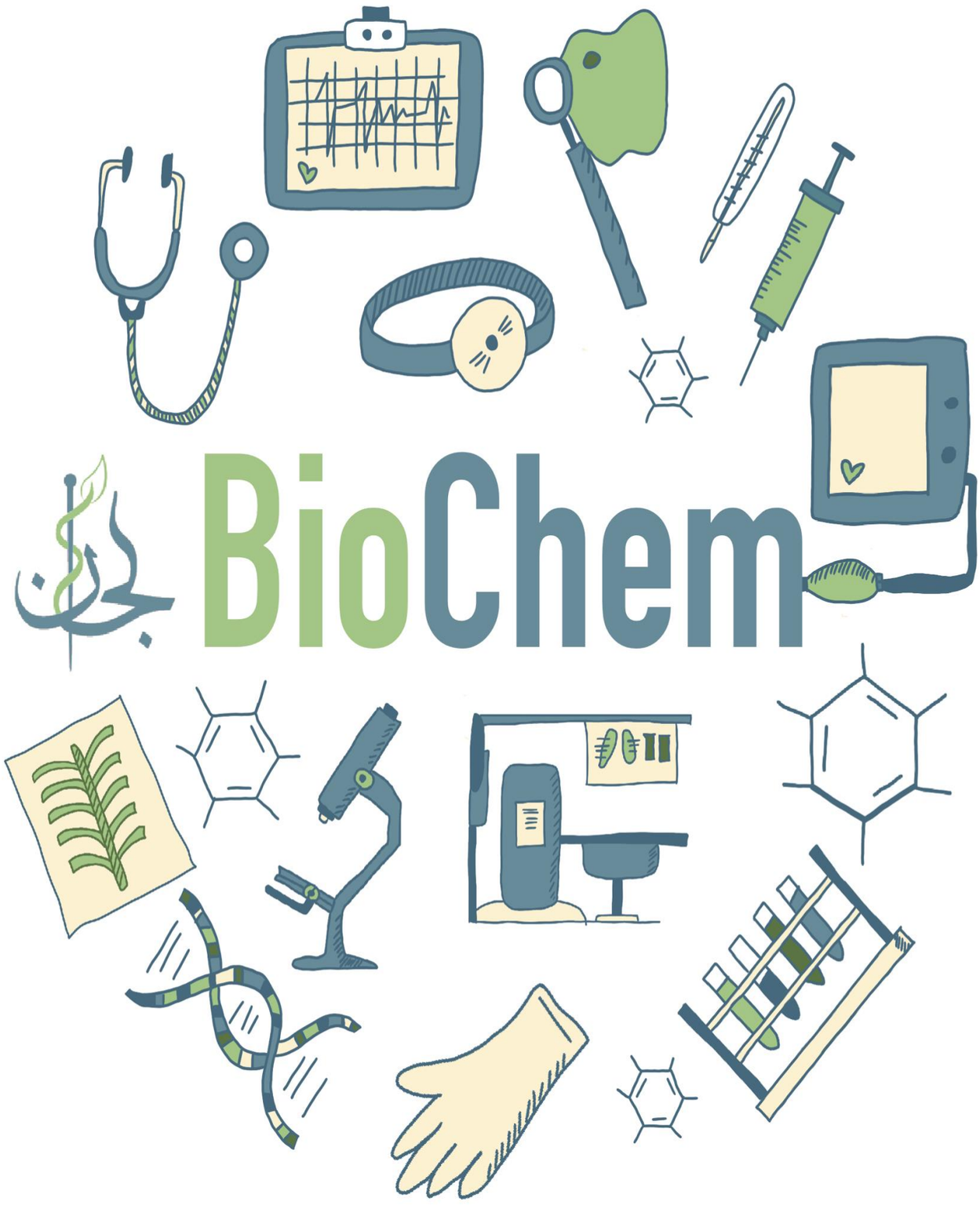


BioChem

Writer: Sara Abdalla

Science: Sema Waleed

Grammar: Jawad Alrawabdeh

Mamoun Ahram

Doctor:

■ Enzyme-based molecular techniques:

1) DNA Sequencing:

DNA sequencing is the process of determining the exact order of

nucleotides in a piece of DNA or in a whole genome.

It's like the process of reading the letters in a book, it is still meaningless, unless you know what the codes are (translation), you can change these letter into meaningful words!

– By knowing the DNA sequence we can identify a lot of things, so what is the importance of DNA sequencing?

1) Identification of genes and their localization, *as if you are identifying words and where they are located in a book.* After knowing the DNA sequence, you will be able to pinpoint that this is a sequence of nucleotides that represents the beginning or the end of a gene and then identifying where the exons and introns are.

2) Identification of protein structure and function, by knowing the sequence of a gene, you can identify the sequence of amino acids of a protein and therefore, you can predict the structure and function of this protein.

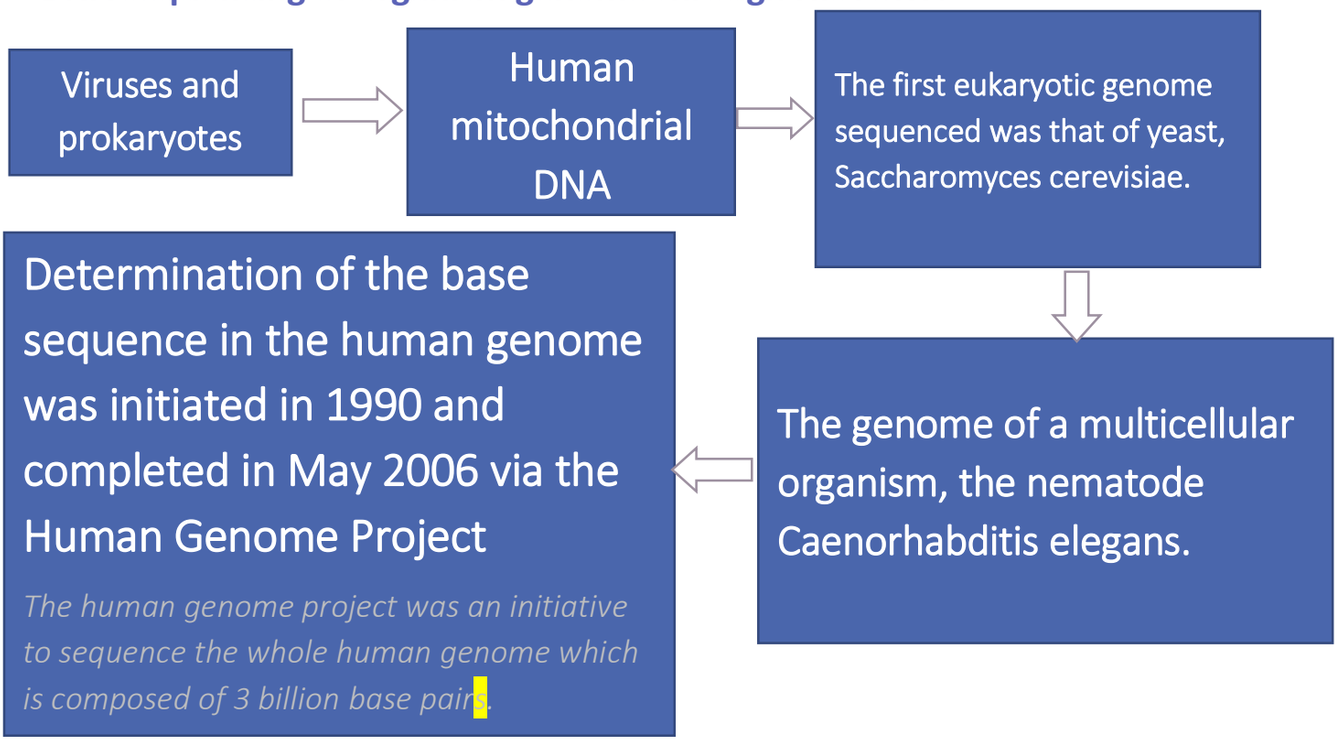
3) Identification of DNA mutations, by knowing what the normal sequence of a gene is, whether it is a coding region, a promoter or an enhancer or whatever, you can identify the mutation, and whether it will cause the production of a defective protein or not which might cause a certain disease.

4) Genetic variations among individuals in health and disease, we are all similar by 99.9%, but we still have 0.1% we are different at, for example: ***single nucleotide polymorphisms*** that increases variation among humans, and that would affect health and disease,

5) Prediction of disease-susceptibility and treatment efficiency, we are able to know the susceptibility (قابلية) to develop a certain condition whether it was cancer or heart disease or whatever.

6) Evolutionary conservation among organisms, DNA sequencing is also important in evolutionary sciences, using relationships and connections between different organisms we can learn about human migration throughout history.

- DNA sequencing of organism genome through HISTORY:



SPECIES	BASE PAIRS (estimated)	GENES (estimated)	CHROMOSOMES
Human (*Homo sapiens*)	3.2 billion	~ 25,000	46
Mouse (*Mus musculus*)	2.6 billion	~ 25,000	40
Fruit Fly (*Drosophilia melanogaster*)	137 million	13,000	8
Roundworm (*Caenorhabditis elegans*)	97 million	19,000	12
Yeast (*Saccharomyces cerevisia*)	12.1 million	6,000	32
Bacteria (Escherichia coli)	4.6 million	3,200	1
Bacteria (H. influenzae)	1.8 million	1,700	1

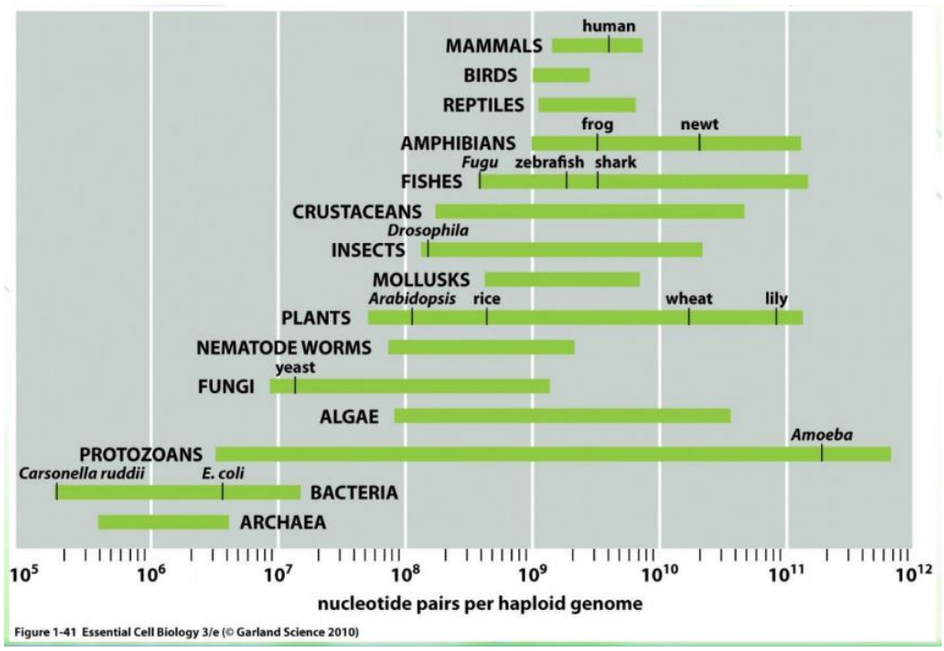
- This figure shows some differences in our genomes compared to other different organisms such as, mice and fruit flies.

NOTICE THAT:

1) There is little difference between human genome and the mouse genome, especially the number of genes. They are really similar to each other (it's not 25,000 anymore, it's about 20,500)

2) In Roundworm (Caenorhabditis elegans), which was used a lot in research, the number of genes is 19,000, so the difference between this genome and the human genome is really nothing, yet we are completely different!

DON'T MEMORIZE.



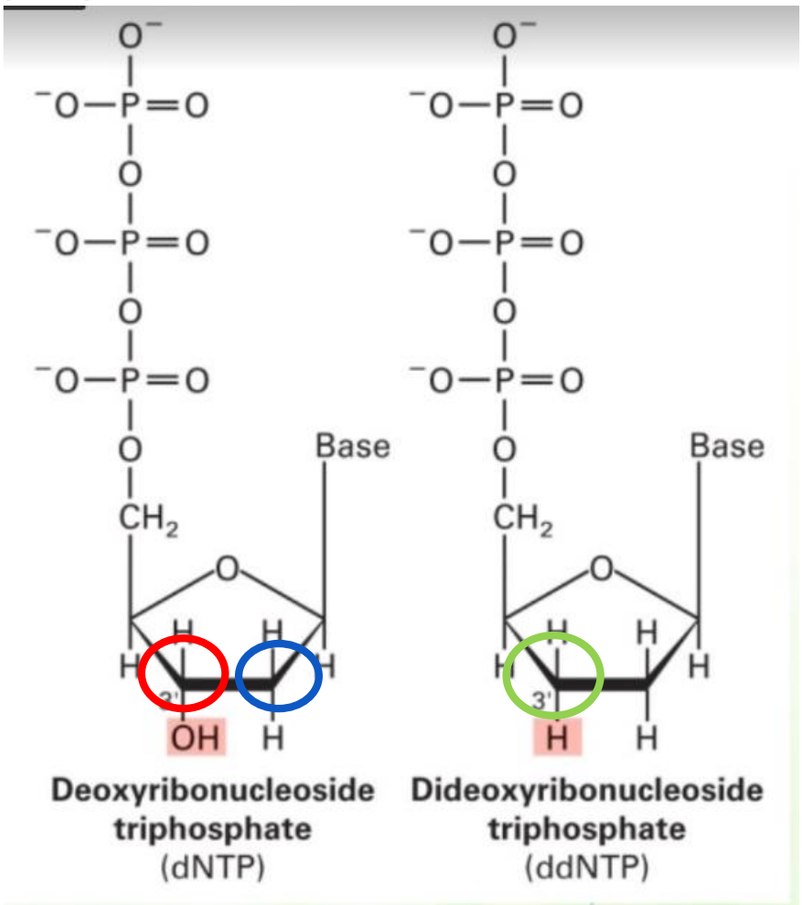
-This figure shows differences in the size of genome, it is smaller in Archaea and Bacteria compared to humans, notice the diversity in Protozoans if you look at different plants and amphibians, they have larger genomes.

DON'T MEMORIZE.

■ Methods of DNA sequencing:

The most popular method is based on **premature termination of DNA synthesis** by **dideoxyribonucleoside triphosphate (ddNTP),** so let's know more about it.

- as we know the normal substrate of the DNA polymerase (during the synthesis of DNA) is **deoxyribonucleoside triphosphate (dNTP)**.



The dNTP has a deoxygenated group on carbon number two (2') (Blue circle), and a carbon with a hydroxyl group which is carbon number 3 (3') (Red circle),hydroxyl group is important in elongating DNA when synthesizing it, as we add the five prime end of the new nucleotide to this end (three prime end) of the last nucleotide.

-Now, by using dideoxyribonucleoside, there is two deoxygenated groups including the one at the 3' end (Green circle), so because this group is no longer a hydroxyl group, no nucleotide can be added to this group, so the DNA synthesis would stop, it terminates.

■ NOW LET'S START DNA SEQUENCING!

1) Tools needed for the process are:

A) We need the DNA template, which we want to sequence.

B) DNA polymerase to synthesis the DNA.

C) The four substrates of the DNA polymerase (the four types of deoxynucleotides) which are : dGTP, dATP, dTTP and dCTP

> ***■ Remember:***
>
> 1) we're not talking about a single DNA molecule rather thousands to millions of the same type, so DNA templates, primers and deoxynucleotides all are in abundant.
>
> 2) synthesis goes from 5' to 3'

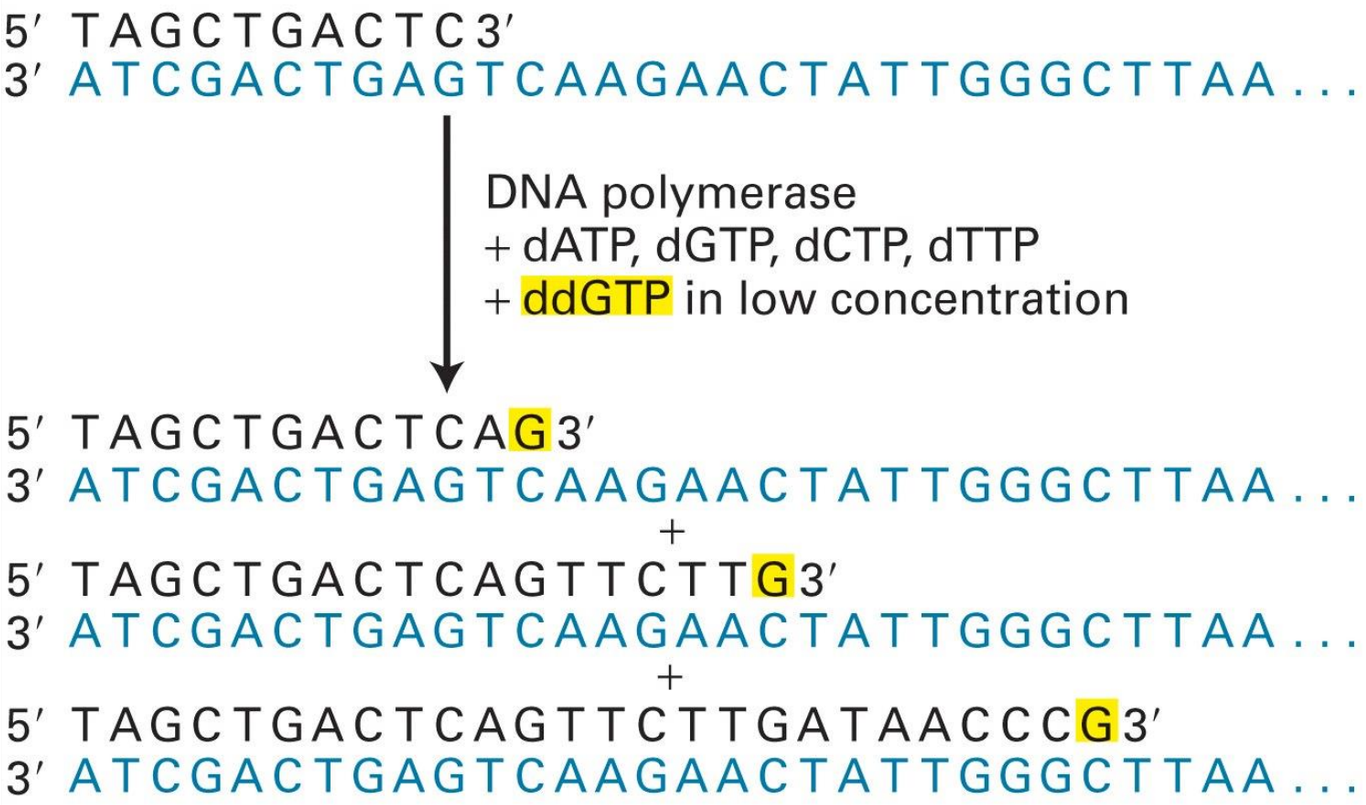
D) A primer that has been labelled with a radioisotope to initiate DNA synthesis, as DNA polymerase can't start DNA synthesis unless there is a starting point. Therefore, I need to have some knowledge about the DNA fragment that I want to determine.

E) the four types of dideoxynucleotides which are: ddGTP, ddATP, ddTTP and ddCTP in low concentration.

**Keep in mind: Incorporation of a dideoxynucleotide stops further DNA synthesis because no 3 hydroxyl group is available for addition of the next nucleotide.*

2) The process:

-Four separate reactions are run, each including same template, same primer, same polymerase, and the same four types of deoxynucleotides, BUT they differ having one dideoxynucleotide (either A, C, G, or T).



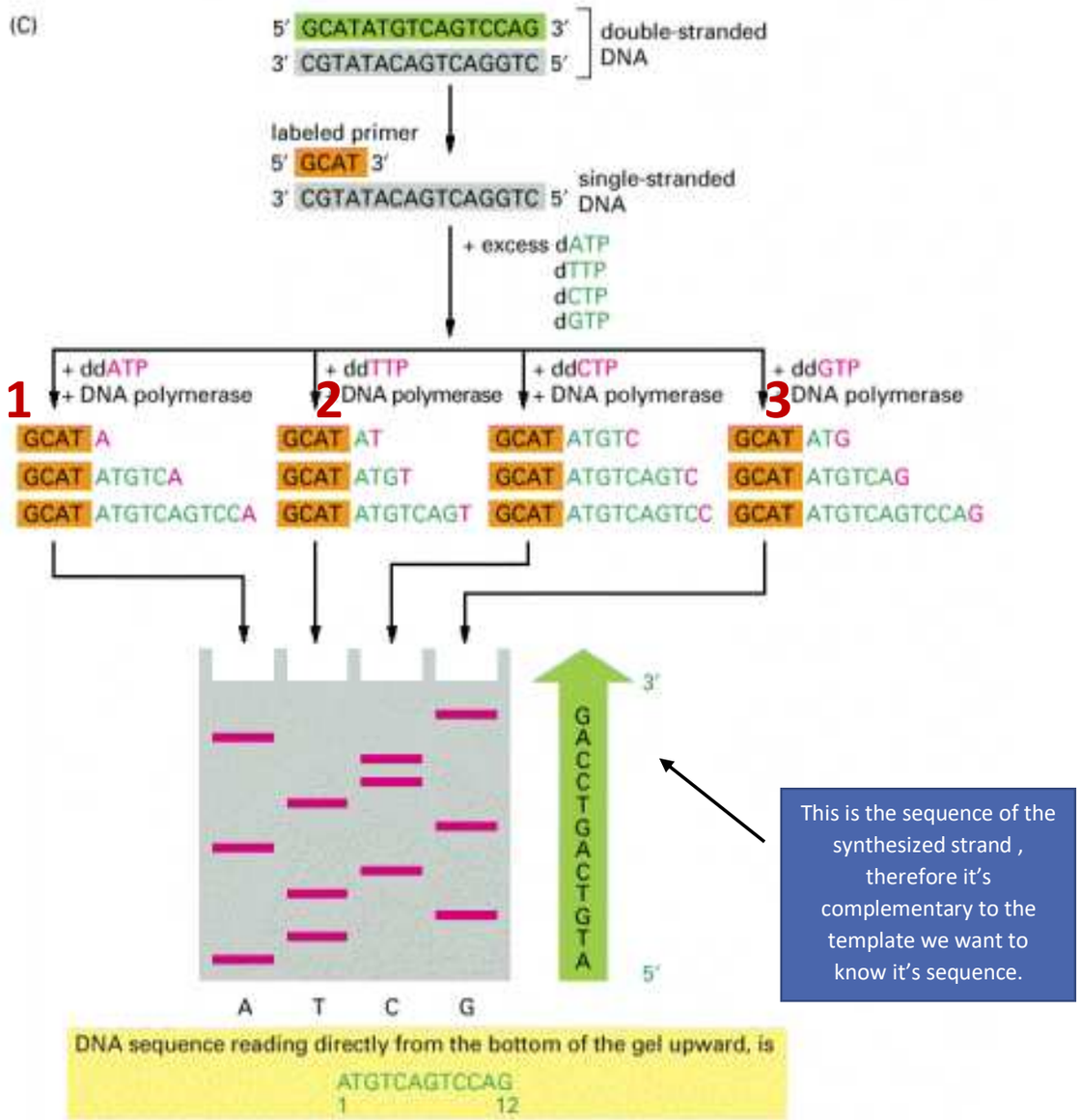
- Imagine that we have 1000 DNA templates for each reaction, each one of them will have a labelled primer attached to it, and we want to know the sequence, so with the presence of the four types of deoxynucleotides and one type of dideoxynucleotides in each reaction, the DNA polymerase will start DNA synthesis.

- As in the figure, the type of dideoxynucleotide used is ddGTP, so whenever there is C in the DNA template there will be a competition between ddGTP and dGTP to be added by the DNA polymerase , and since ddGTP presents in **LOW CONCENTRATION,** there is a good chance for dGTP to be added, so the synthesis will continue, BUT we can't ignore the chance for ddGTP to be added, so if ddGTP is added there will be a termination in the DNA synthesis, so as I have 1000 DNA template let's say that 100 will be terminated, the other 900 will continue DNA synthesis, until DNA polymerase reaches another C, there will be another competition between ddGTP and dGTP , therefore another 100 DNA templates will be terminated, and 800 will continue and so on,

- NOW, A series of labeled DNA molecules are generated, each terminated by the dideoxynucleotide in each reaction, so we have DNA fragments with different sizes from each reaction (depends on where the ddNTP is added), in the fragments that are produced from the reaction where ddGTP is used, the last nucleotide will be G (in the synthesized strand) , and in the fragments that are produced from the reaction where ddATP is used, the last nucleotide will be A and the same concept for the reactions where ddTTP and ddCTP are used.

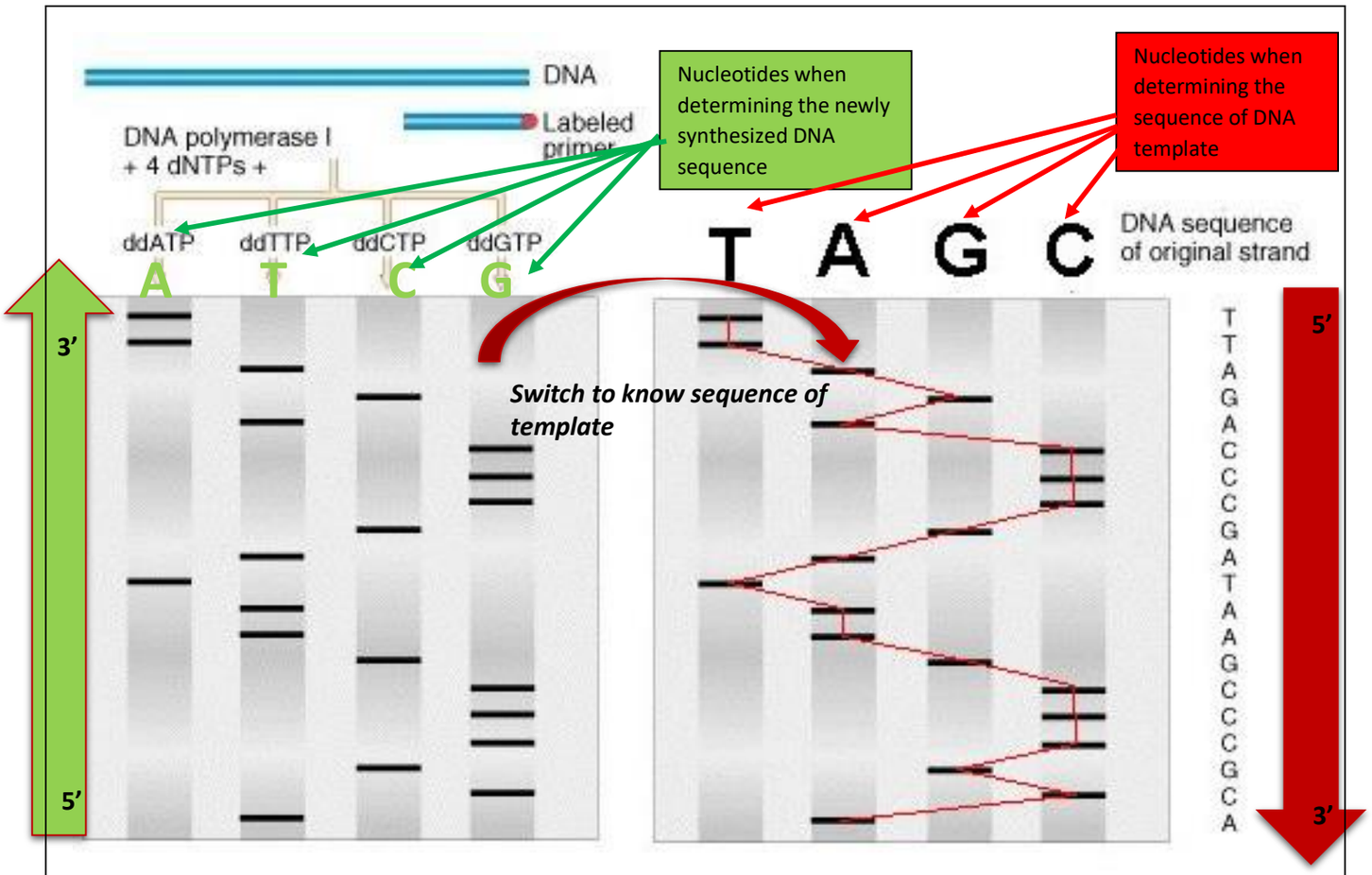
- These fragments of DNA are then separated according to size by gel electrophoresis and detected by exposure of the gel to X-ray film.

The size of each fragment is determined by its terminal dideoxynucleotide, so the DNA sequence corresponds to the order of fragments read from the gel.

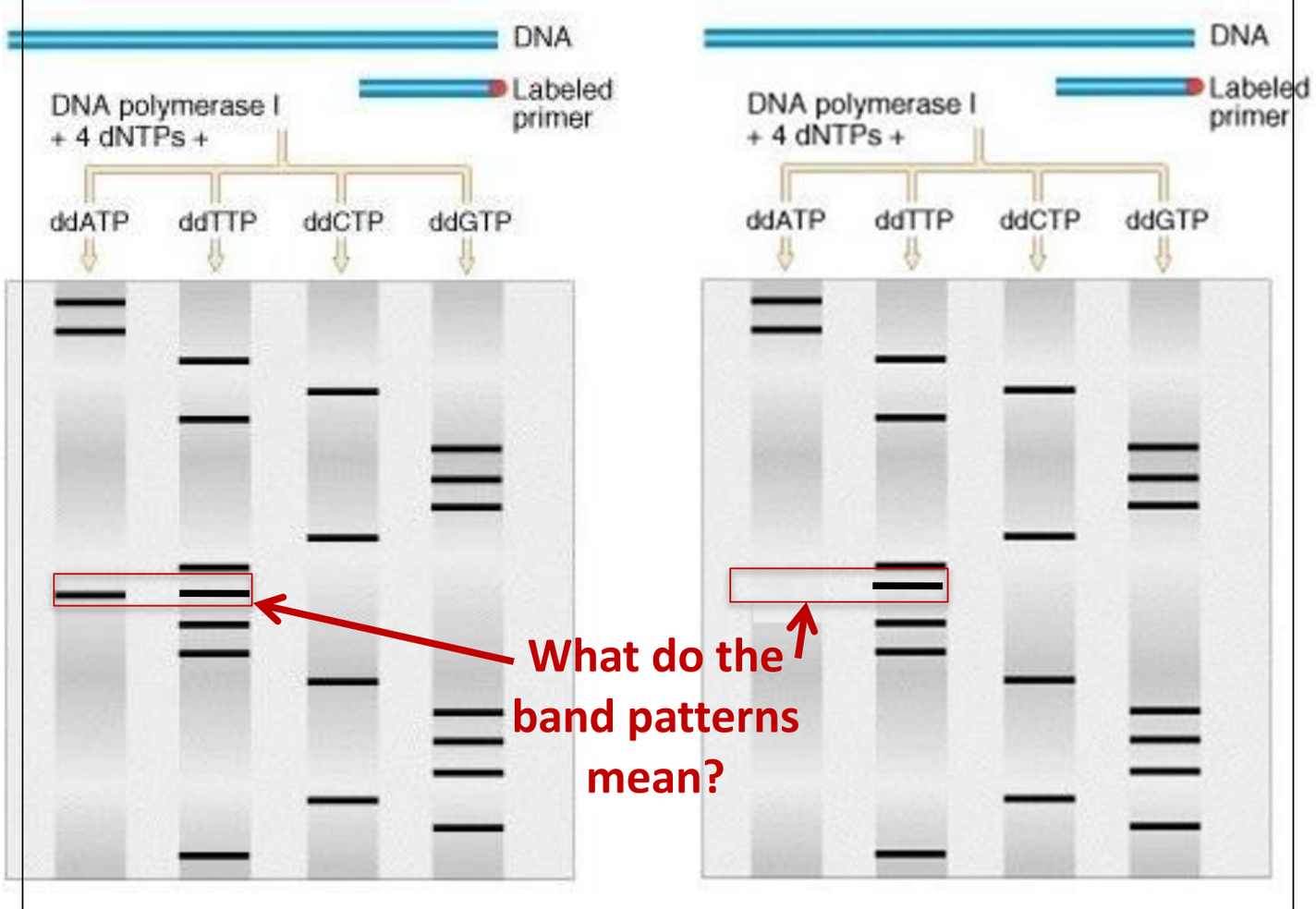


- Each fragment differs from the second fragment in size by one nucleotide only so the separation has very high resolution, it depends on only one nucleotide. For example, fragment 1 is smaller than fragment 2 by one nucleotide, and fragment 2 is smaller than fragment 3 by one nucleotide.

- As in the figure, we can determine that the first nucleotide is A **in the newly synthesized** sequence (from 5' because synthesis starts from 5' to 3'), because in the smallest fragment (passed longest distance) synthesis is terminated by ddATP, and with going from the fragment that passed the longest distance (smallest) to the fragment that passed the shortest distance (largest) we can determine the sequence of the newly synthesized strand from 5' to 3'.



The sequence of the newly synthesized DNA will be determined from 5' to 3'(green arrow), so if you want to know the sequence of the DNA template , you have to flip because it's anti-parallel, to get the sequence from 5' to 3' you start from up (red arrow) , and also keep in mind that it is complementary to the newly synthesized DNA, therefore you will consider column one as T not A when determining the sequence of DNA template.



What do the band patterns mean?

1) Lets say that I have a gel, as in the first figure, and I have two bands at the same exact position, so they have the same exact length, how is that possible?

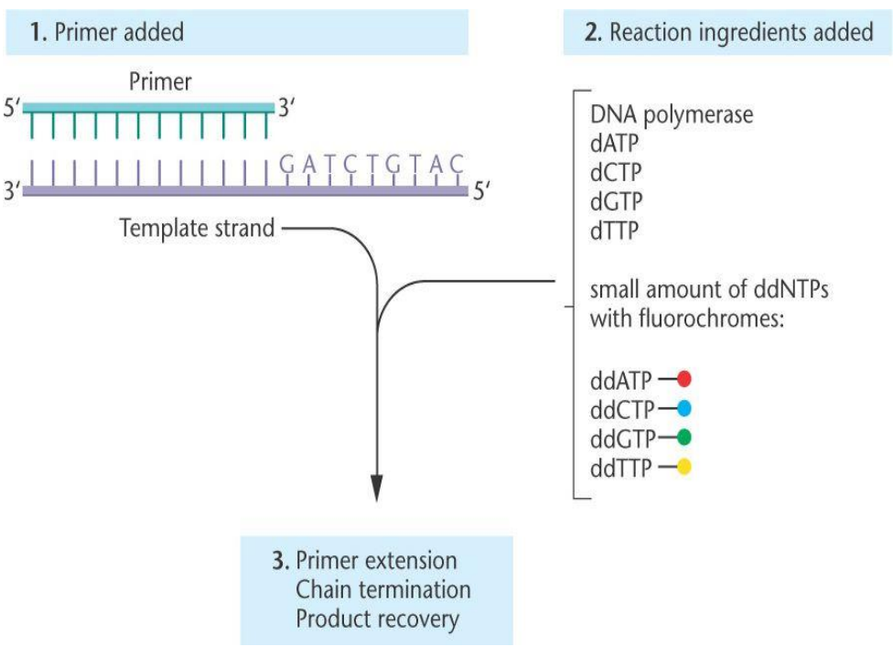
Our cells are diploid, we have two copies of every chromosome one from the father and one from the mother, so there is possibly a variation by one nucleotide between the father's DNA and the mother's DNA in that particular DNA we are sequencing, so let's say that DNA polymerase is reading both DNA in that individual until it reaches a variation between the two molecules for example, it may be A in the father's and T in the mother's. so DNA will add T to the father's and A to the mother's, and then will continue with the same exact sequence, because there is only variation in one nucleotide, so we are going to have DNA fragments with the same length as a result of what we call **polymorphisim.**

Also , it might be a mutation in one of the alleles mother's or father's.

2) Look at the second figure, let's say that normally there should be a band in this position, but it's not here, it appears in the other lane, that means that normally there is A in each allele, but there is a mutation in both chromosomes of the mother and the father, or maybe variation, so instead of having A there is a T.

Fluoresence- based DNA sequencing:

- In the first method of DNA sequencing we were using radioactive primers, and radioactivity is really harmful for our cells and our DNA molecules, it's also expensive, so scientists decided to use something less harmful such as fluoresence, as it is sensetive and not dangerous, and since they were using fluoresence they chose to let the computer read the gels instead of reading it manually, so instead of labelling primers they decided to label the nucleotides themselves, and each nucleotide will give a certain color.



-Look at the figure, reactions include the four deoxynucleotides plus THE FOUR dideoxynucleotides in the same reaction, since I can differentiate each one from the other ,as each ddNTP is labeled with a unique fluorescent tag.

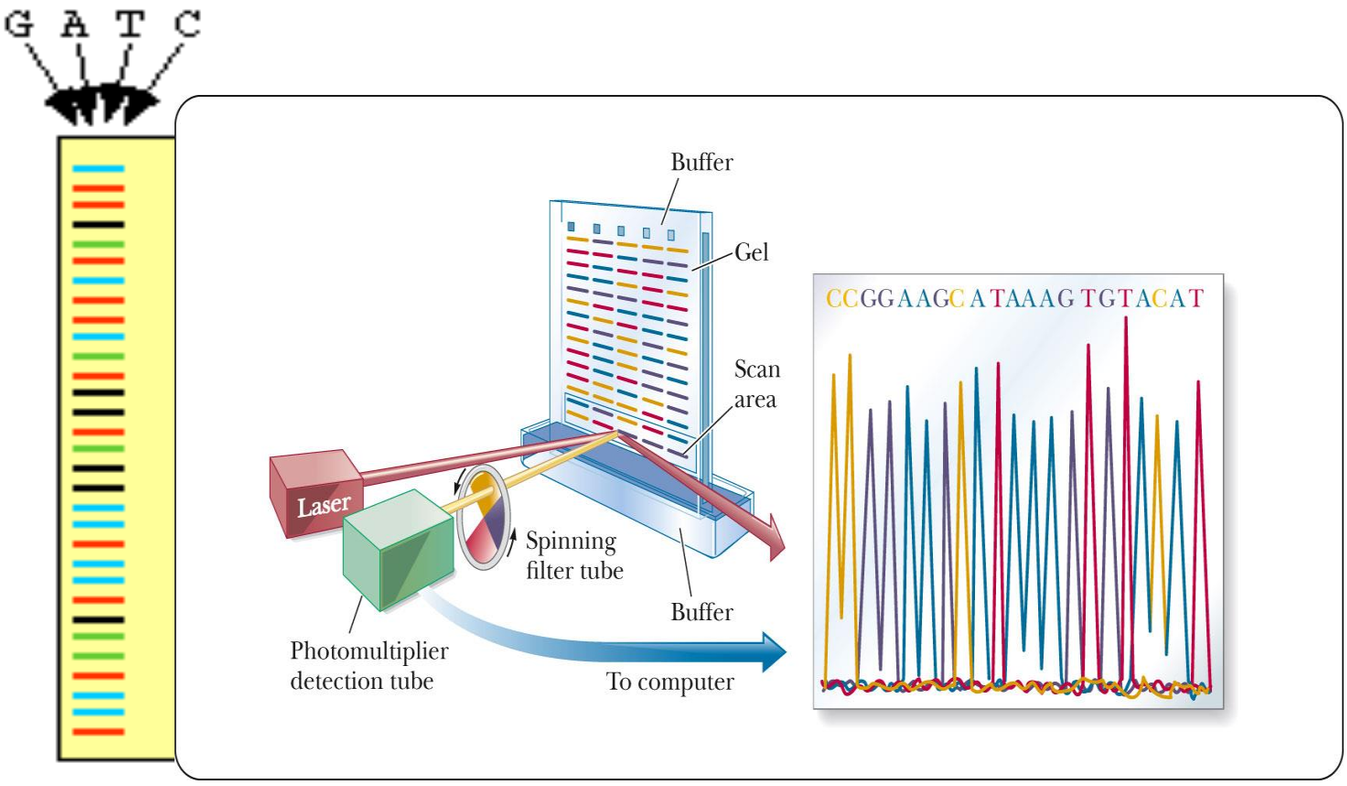
-The main differences between the two methods are:

1) There will be termination at every single nucleotide, since all the dideoxynucleotides are added together in the same reaction.

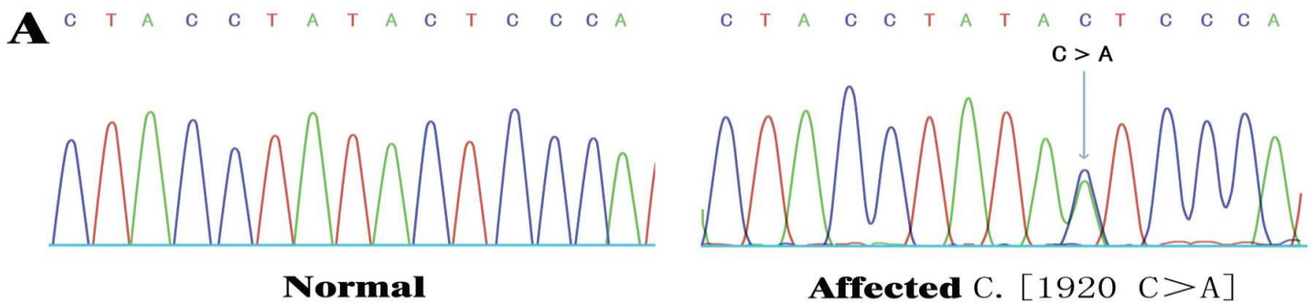
2) There will be a detector analyzing the signal that comes out from each band, which gives a specific color depending on which dideoxynucleotide is added to it and terminated the DNA synthesis, also each band has a specific length.

As in the figure, the computer gives us the signals as peaks

-Refer to the animation videos to understand better.



-What if there is a variation or a mutation, how would that look like?



- Look at the figure, notice the peaks in the normal condition, and in the affected condition, you can see that in the affected condition there is an overlap, two peaks with similar height (almost the same in terms of amount) and same position (same size) BUT different colors, so at the marked position there is C and A signals one from the father and the other from the mother.

If there is a mutation in one allele and the other one is normal the result will be the same, as if there is variation.

NOTE(not mentioned in the lecture):

The height of the peak represents the amount of fragments in each band. So Look at the figure, notice the height in the normal condition (where all the molecules in the band is terminated by C (both mother and father give C)) is higher than the two heights of the two peaks in the affected condition where there is variation, it seems that the height is divided and gave two peaks with different colors, because the amount is divided , so each peak represent the father or the mother therefore, the two peaks are lower in height than the one peak in the normal condition.

I will ask the doctor about the accuracy of this info.

BUT..

- What if there is a mutation in both alleles or genes?, if ,for example there is C in the first allele and C in the second allele instead of G and G, normally we will have a single peak with a color representing G, if there is a mutation we will have a single peak as will but representing the color of C

■Next-generation sequencing:

All the previous techniques are laborious, so scientists keep on thinking of improving technologies, since sequencing is done nucleotide by nucleotide, they thought of an easier way, so they came up with different technologies, we will take about one of them, collectively, they are called next-generation sequencing(NGS) a really fast sequencing of DNA.

The process is :

We take a whole genome/DNA from a cell, and then we cut it into different, random fragments, no endonuclease is used, since it cuts at a certain position, it's completely random fragmentation, then DNA adapters are added to the end of these fragments, these adapters are added for two purposes:

1) To allow these DNA fragment to attach to a platform, a solid surface.

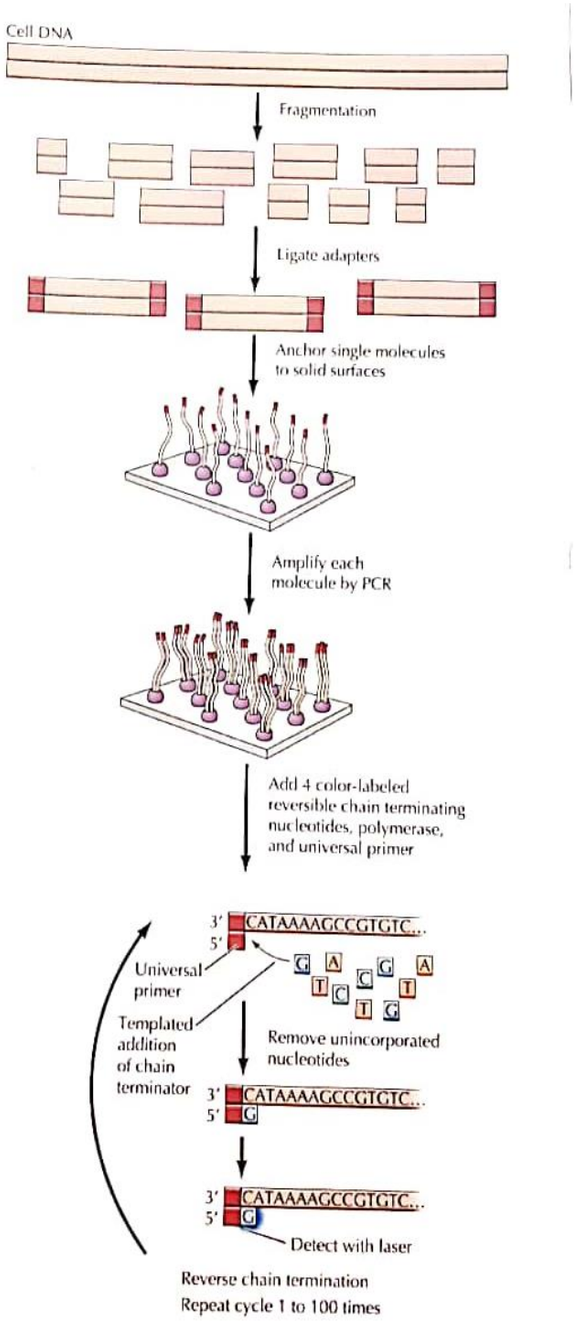
2) They act as binding sites of primers.

Then DNA synthesis starts special nucleotides are added, these nucleotides have two properties:

1) Once one of them is added the next nucleotide will not be added, unless the first nucleotide is activated by being detected with a special camera. (they are NOT dideoxynucleotides, they are special nucleotides)

**These nucleotides act as dideoxynucleotides (don't allow addition of the next nucleotide),but they are special that after modifing them (when we detect them) they will be activated and allow the addition of the next nucleotide)

2) They can fluoresce each type has different color,give a signal which is detected by a special camera.



So the steps are:

1)Cellular DNA is fragmented.

2)DNA adapters are added to ends of each DNA fragment.

3)Each DNA fragment is attached to a solid surface and amplified like PCR using primers that anneal to the adapter sequences.

4)Four-color nucleotides with terminating ends are added.

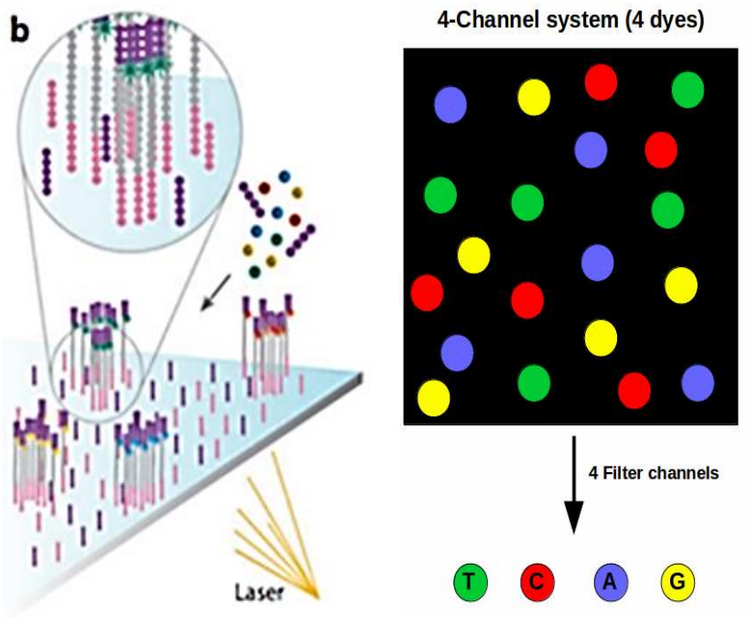
5) A single nucleotide is incorporated and unincorporated nucleotides are removed.

The incorporated nucleotide is modified for two reasons:

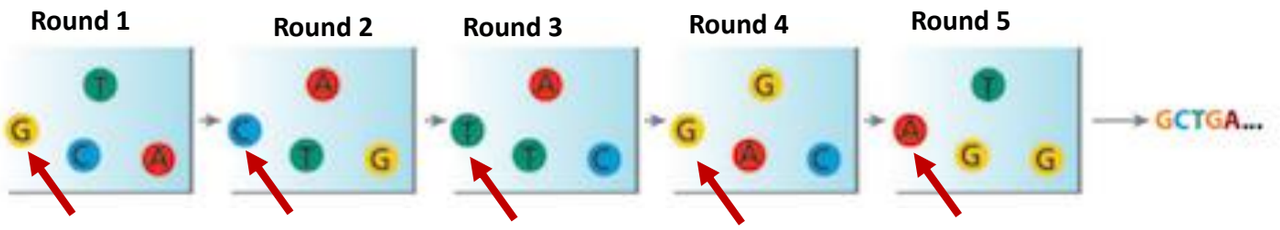
1) It is activated and detected by a special camera.

2)A new nucleotide can be added to it.

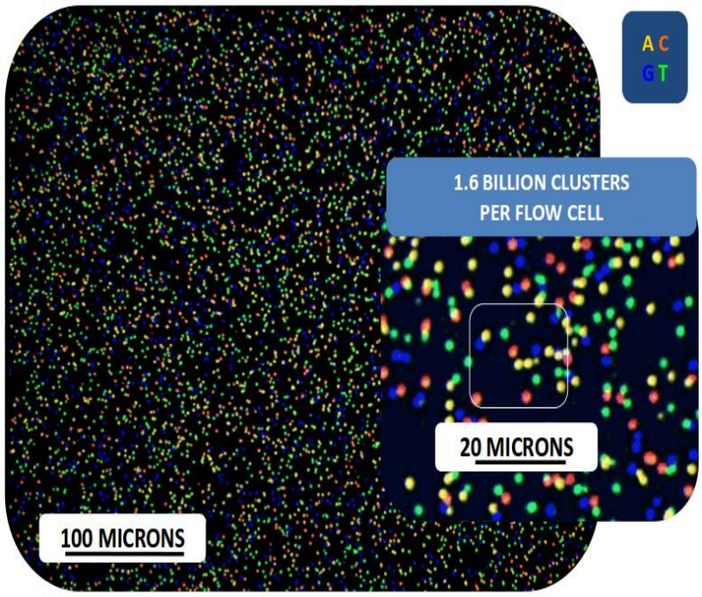
6)The cycle is repeated.



- We are talking about billions of fragments added, so they will be like the first figure, each spot from these different spots represents one nucleotide detected from one fragment with a specific color, then, after detecting(round one), this special nucleotide will be activated, and another nucleotide will be added, the color of this nucleotide will be detected too (round two) and so on.



- Look at the second figure, in round one, the yellow spot represents the first nucleotide in one of the fragments, and the blue spot represent the first nucleotide of another fragment, in round two, the spots will change, the colors will represent the second nucleotide in each of the fragments, this process is very fast because these are enzymatic reactions which are really fast.



- All of these spots represent different fragments that are all synthesized at the same time, each one produces a different color in less than microseconds, the camera is really fast it detects all of these colors at the same time, this is achieved really quickly, by which the Human Genome Project was achieved.

Note: we can have sequencing of the both strands of the DNA at the same time.

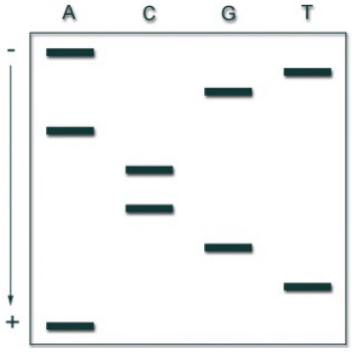
- Don't forget to watch the video, details are not required from it.
https://www.youtube.com/watch?v=womKfikWlxM

BEST WISHES ☺

SHORT QUIZ

TEST BANK

1. Starting from the sequencing primer, what is the sequence of the original DNA sample?



a) ATGACCGTA
b) TACTGGCAT
c) ATGCCAGTA
d) UACGGUCUA
e) TACGGTCAT

2. If we have 2 dATPs, 1 dCTP, 1 ddCTP, and 2 ddGTPs in one reaction tube, which of the following strands could be produced from a sample containing the following template strand: 5' GCTTGGCTTAACCAGATATTCCACTG 3' with the following primer: 5' CAGTGGAATATCTGGTT 3'?

a) 5' CAGTGGAATATCTGGTTAAG 3'
b) 5' CAGTGGAATATCTGGTTAAGCC 3'
c) 5' CAGTGGAATATCTGGTTAAGCCAA 3' d) Just 1 and 2
e) 1,2, and 3 are possible

3. Based on this figure, the most likely error is:

a) The scientist forgot to add dNTPs to one of the reaction tubes
b) The scientist did not fluorescently/radioactively label one of the ddNTPs
c) The scientist mixed up the cathode and the anode during gel electrophoresis
d) The scientist added two different ddNTPs to one of the reaction tubes

4. Based on this figure, we can deduce what DNA sequence? (assume the anode is at the bottom)

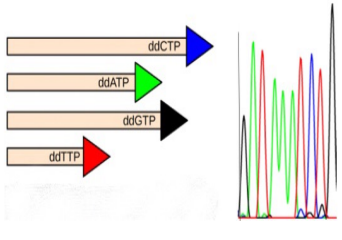
a) 5' CACTCAATGTCATGCTGCAT 3'
b) 5' TACGTCGTACTGTAACTCAC 3'
c) 5' GAGTGAATGTGATGGTGGAT 3'
d)5' CACTCAATCTCATCCTCCAT 3'
e) The scientist's error renders it impossible to determine the DNA sequence

5. If you wish to sequence a long strand of DNA in one round of reactions, you should:

a) Increase the ddNTP/dNTP ratio
b) Decrease the ddNTP/dNTP ratio
c) Use a shorter DNA primer
d) None of the above

6. What is the DNA sequence:

a) GTCTAAATAG
b) CTATTTAGAC
c) GATAAATCTG



ANSWERS

Q1	Q2	Q3	Q4	Q5	Q6
E	A	D	E	B	C