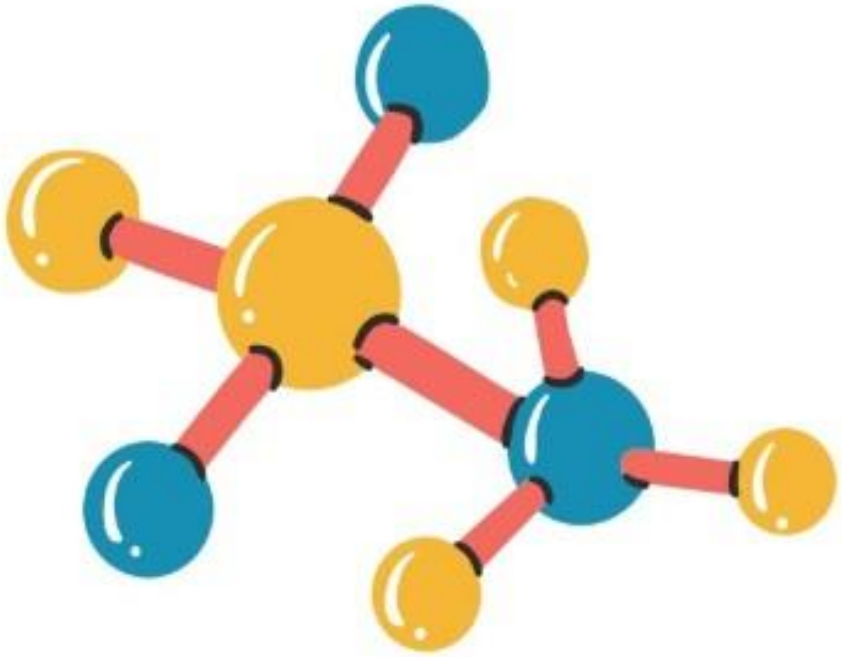
Sheet no.16

Biochemistry



Summer 2022

Writer: Aseel rommaneh,
Corrector:
Doctor : Mamoun ahram

Slides
Doctor`s notes
Writer`s notes

Enzyme-based molecular techniques (part I)

DNA sequencing:

First, we have some definition

-DNA sequence : is the process of knowing the exact order of nucleotide in a genome or DNA fragment

*So basically I give you DNA fragment or whole genome i ask you what is the sequence of this DNA , so you take it from 5` to 3` and you start reading the DNA sequence like AGCTAAT.., so that is DNA sequence

The purpose of knowing DNA sequence is:

1) Identify genes and where they are localized

*You tell me that gene X is located on this arm of chromosome, near telomere, near centromere, in the middle and there are certain number that indicate where gene is localized and it is next to this gene ..so on

2) Identify protein structure and function

*If we identify a gene we identify a codon and translate the codon to amino acid so we would able to know the sequence of the protein , by knowing the sequence of protein we can predict what the structure is it, and there are recently artificial intelligence and bioinformatics tool that are successfully with high accuracy predict the protein structure , by the knowing protein structure and the sequence of protein we can predict the functionof this protein , as well as know what protein it can interact with.

3)identify DNA mutation

*We have databases right now for what normal human genome should look like andwe can compare sequence of our unknown DNA or DNA with certain disease we can compare it to this database and we can pinpoint exactly where mutation occurs andhow it is related to certain
Disease.

4) sort of like elucidate clarify the genetic variation among individual among population in health and disease

*By knowing DNA sequence of an individual we can know how variable anotherindividuals on to other individual.

5) predict the disease susceptibility of an individual and treatment

*by knowing DNA sequence we know there is 50 percent chance that an individual might get a certain disease, pay attention we don't take medication because you have genetic variant you have this nucleotide in this gene and that makes you slow metabolizers for medication , so we will be able to help individual by letting him to know to take care of himself and we can determine the best treatment by knowing sequence of cancer forexample we give him the best solution.

6) determine how organism are related to each other in terms of evolutionary

*We able to compare human gene with mouse or dog this help us to understand our cell , our disease and how cell function normally or there if is a mutation because we use animal as model system for human.

These are benefits of knowing DNA sequence Now , the brief history of DNA sequence

1) We started with simple organism , simple genome like viral genome , bacterial genomebecause they are small and easy to handle.
2) Human mitochondrial DNA.
3) Simple eukaryotes the yeast cell the model system , Saccharomyces cerevisiae
4) Multi-cellular organisms like nematodes Caenorhabditis elegans
5) Human genome sequence and it hasn't complete yet , we still have the final reasonsthe final piece of Y-chromosome

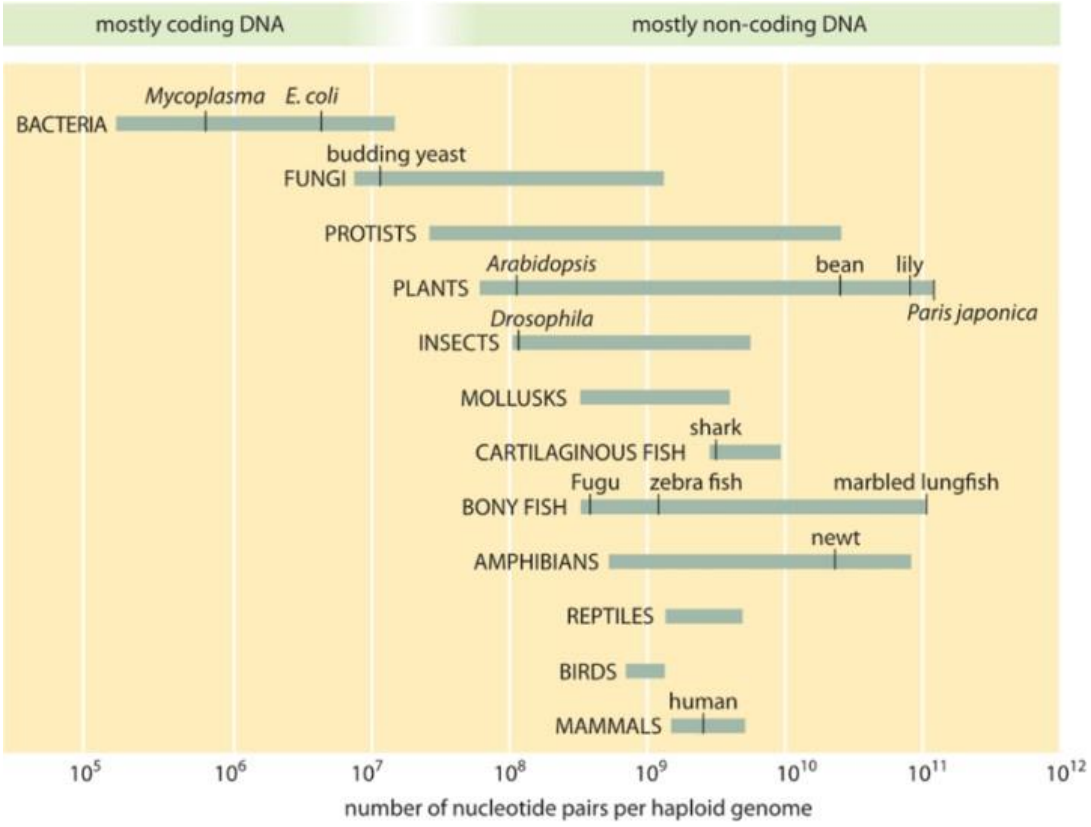
This table show comparison of genome of different organism and number of protein coding gene and number of chromosomes

You can look to it , its not for memorize

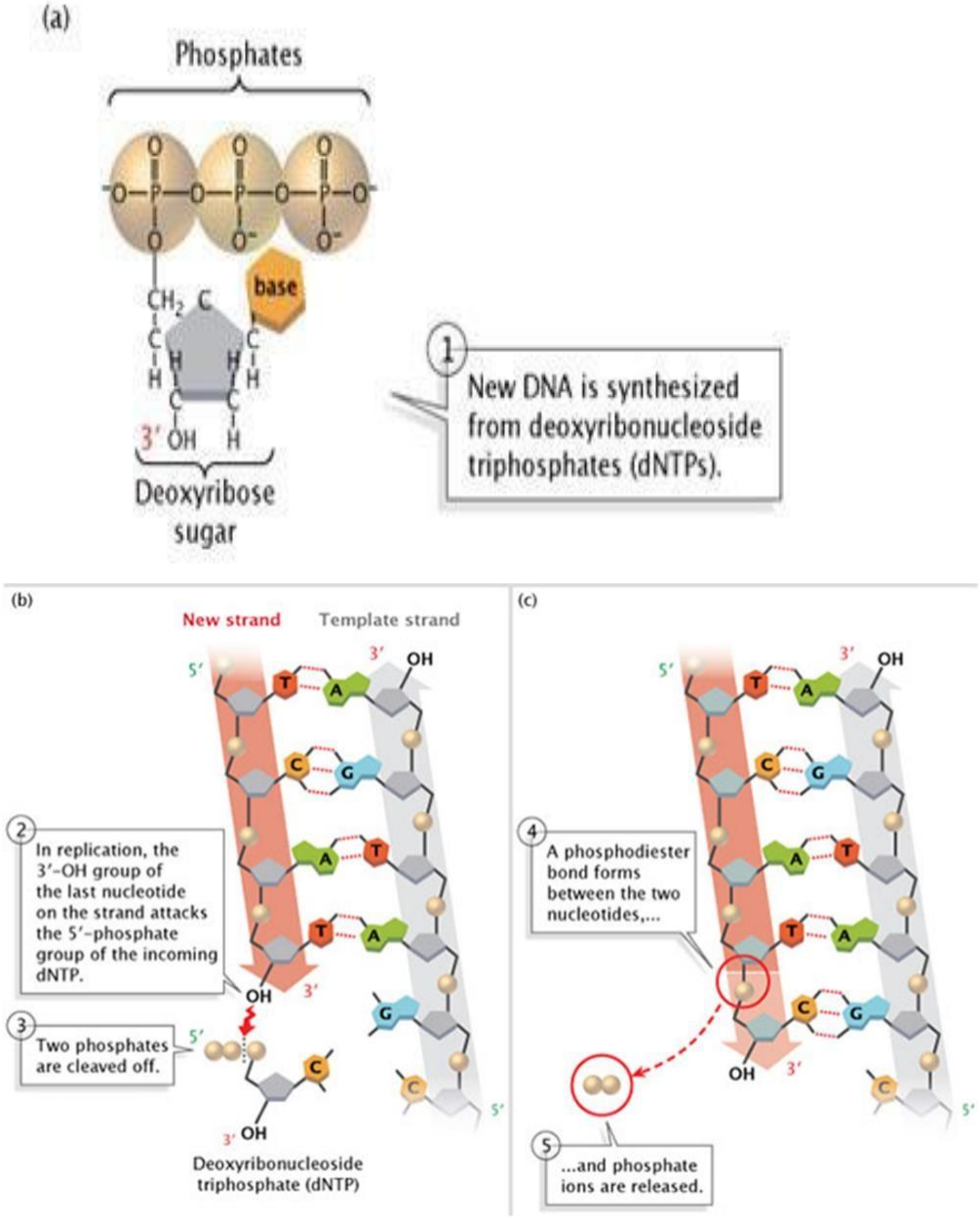
organism	genome size (base pairs)	protein coding genes	number of chromosomes
model organisms			
model bacteria *E. coli*	4.6 Mbp	4,300	1
budding yeast *S. cerevisiae*	12 Mbp	6,600	16
amoeba *D. discoideum*	34 Mbp	13,000	6
nematode *C. elegans*	100 Mbp	20,000	12 (2n)
fruit fly *D. melanogaster*	140 Mbp	14,000	8 (2n)
model plant *A. thaliana*	140 Mbp	27,000	10 (2n)
mouse *M. musculus*	2.8 Gbp	20,000	40 (2n)
human *H. sapiens*	3.2 Gbp	21,000	46 (2n)
viruses			
hepatitis D virus (smallest known animal RNA virus)	1.7 Kb	1	ssRNA
HIV-1	9.7 kbp	9	2 ssRNA (2n)
influenza A	14 kbp	11	8 ssRNA
bacteriophage λ	49 kbp	66	1 dsDNA
organelles			
mitochondria - *H. sapiens*	16.8 kbp	13 (+22 tRNA +2 rRNA)	1
chloroplast – *A. thaliana*	150 kbp	100	1
eukaryotes - multicellular			
dog *C. familiaris*	2.4 Gbp	19,000	40
chimpanzee *P. troglodytes*	3.3 Gbp	19,000	48 (2n)

This table show comparison of genome of
different organism starting with bacteria
ending with mammals including human , you can sea variation in
sizesin different species and it related to complexity but not100%

Not for memorizing too



This is a nucleotide and it has sugar which is deoxyribonucleotide because it is missinghydroxyl group , at carbon n3 we have hydroxyl group attached to it , at carbon n5 attached to triphosphate group



where ever DNA synthesize you have attachment with hydroxyl group at 3`, you have phosphodiester bond formation between carbon n3 and carbon n5 , note the energy is needed and it comes of release two phosphate group and they formation of phosphodiester between phosphate and hydroxyl group and it continues from3`

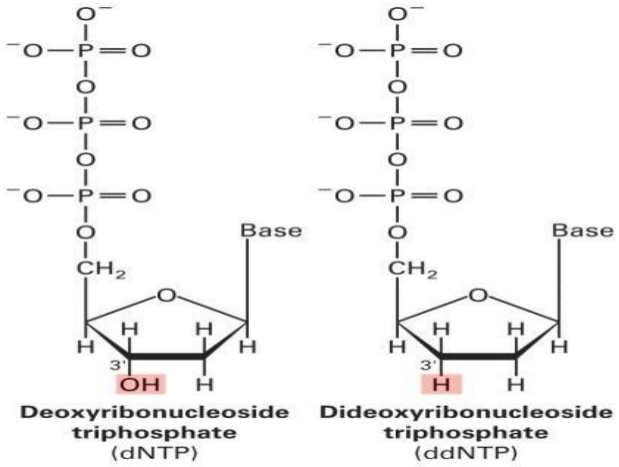
In order to synthesize DNA what we need?

1- Triphosphate nucleotide

2- Hydroxyl group

3- 3` end

Lets talk about technique of DNA sequence



dideoxy nucleotide is similar to deoxynucleotide that is have again the sugar , the base, 3`, phosphate group and youhave 3` with hydroxyl group

in the picture, this is the dideoxy because it has a deoxy carbon that is missing hydroxyl group (oxygen) , this mean that where ever this is added to DNA no other nucleotide canadded to 3` and no phosphodiester bond will be form because there is no hydroxyl group ,there is no reactive group , so DNA synthesis as a result stops at this point so this would be very last nucleotide that is added to DNA andthat is used for DNA sequence

The process

In order to sequence DNA what we need?

1) Primer:

a) just 1 because we are going to only sequence 1 strand

b) is labeled with radioisotope like radioactive phosphorus

c) we need it because DNA polymerase cannot start DNA synthesis denovo that isfrom scratch right

2) DNA polymerase

3) substrate for DNA polymerase that is four deoxynucleotides (A,C,G,T) so we are goingto have four reaction

In these four tube we have to have (template to be sequence, DNA polymerase, 4substrate , single primer, 1 deoxynucleotide in each tube)

The idea here that is DNA polymerase wouldstart synthesizing DNA but then if it add deoxyribonucleotide will be continue , but if we add dideoxy ribonucleotide synthesis it stops because no other nucleotide can be added to deoxy 3` of DNA

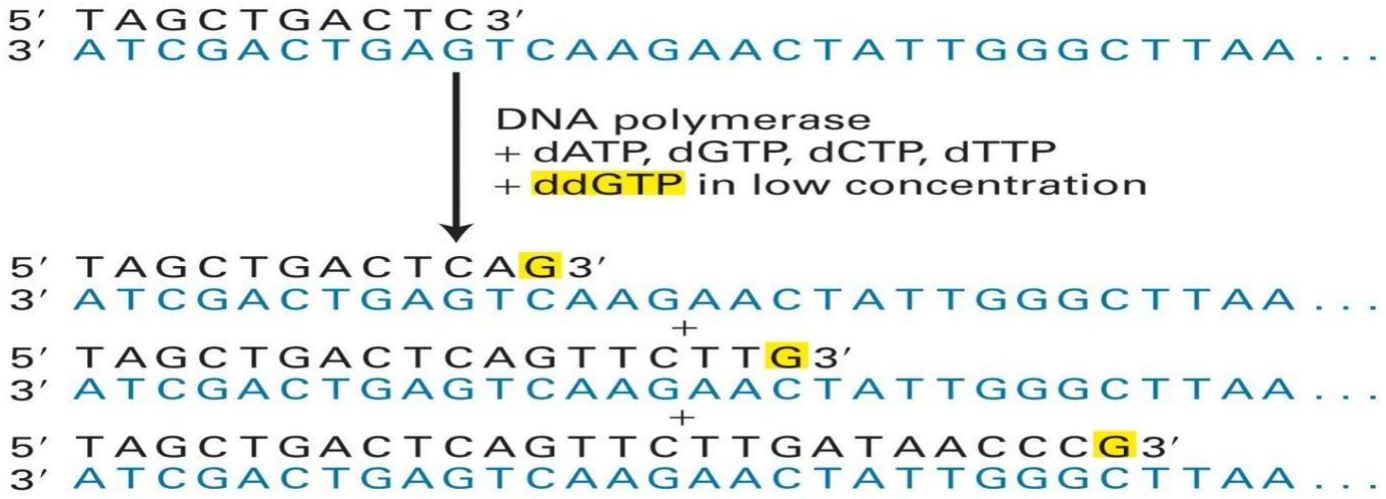
Generation of fragment

What we do then?

We separate DNA fragment that are formed from each one of these tubes through gel

electrophoresis so we will have differentfragment and length .

DNA fragment by the way differ by one nucleotide so we are going to have sequenceof different nucleotide by 1 nucleotide then we can express the gel to x-ray film and we detect signal that's if we used radioactivity But no one use radioactivity anymore becauseits risky and there is easier method



we want to know the order of nucleotide , weuse primer we know the sequence of the primer and this primer is labeled with radioactive phosphorus. we add to the each tube DNA polymerase ,substrate , dideoxy G with low concentration

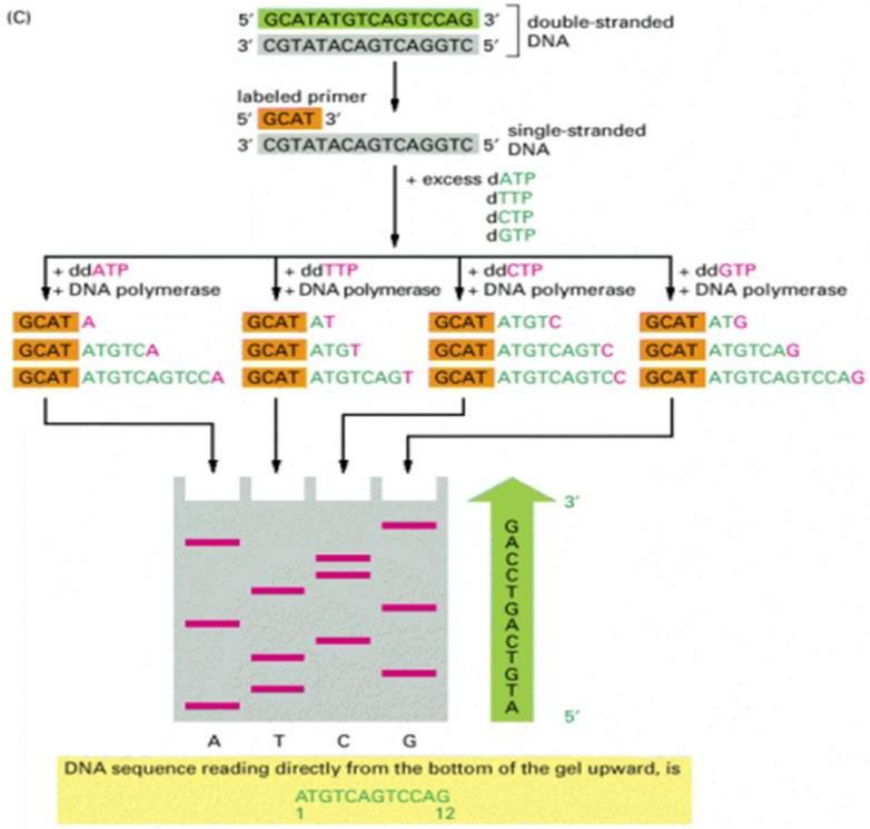
Where ever DNA polymerase see C it add G ,

it has 2 choices either deoxy G or dideoxy G .

The probability is to add deoxy G becausedideoxy exist in low concentration

بتعمل الخيارين لكن بنسب مختلفة

It adds deoxy more so the will continues but if it adds dideoxy the synthesis will stop



We have 4 tubes in each of them we havedouble stranded DNA that we know the sequence , labeled primer ,substrate for enzyme, the DNA polymerase enzyme.

Again we have(enzyme ,template ,primer, substrate) but for each one we could add dideoxysubstrate.

The primer is labeled so it gives us signal sowe can separate the DNA fragment in gel electrophoresis according to the size.

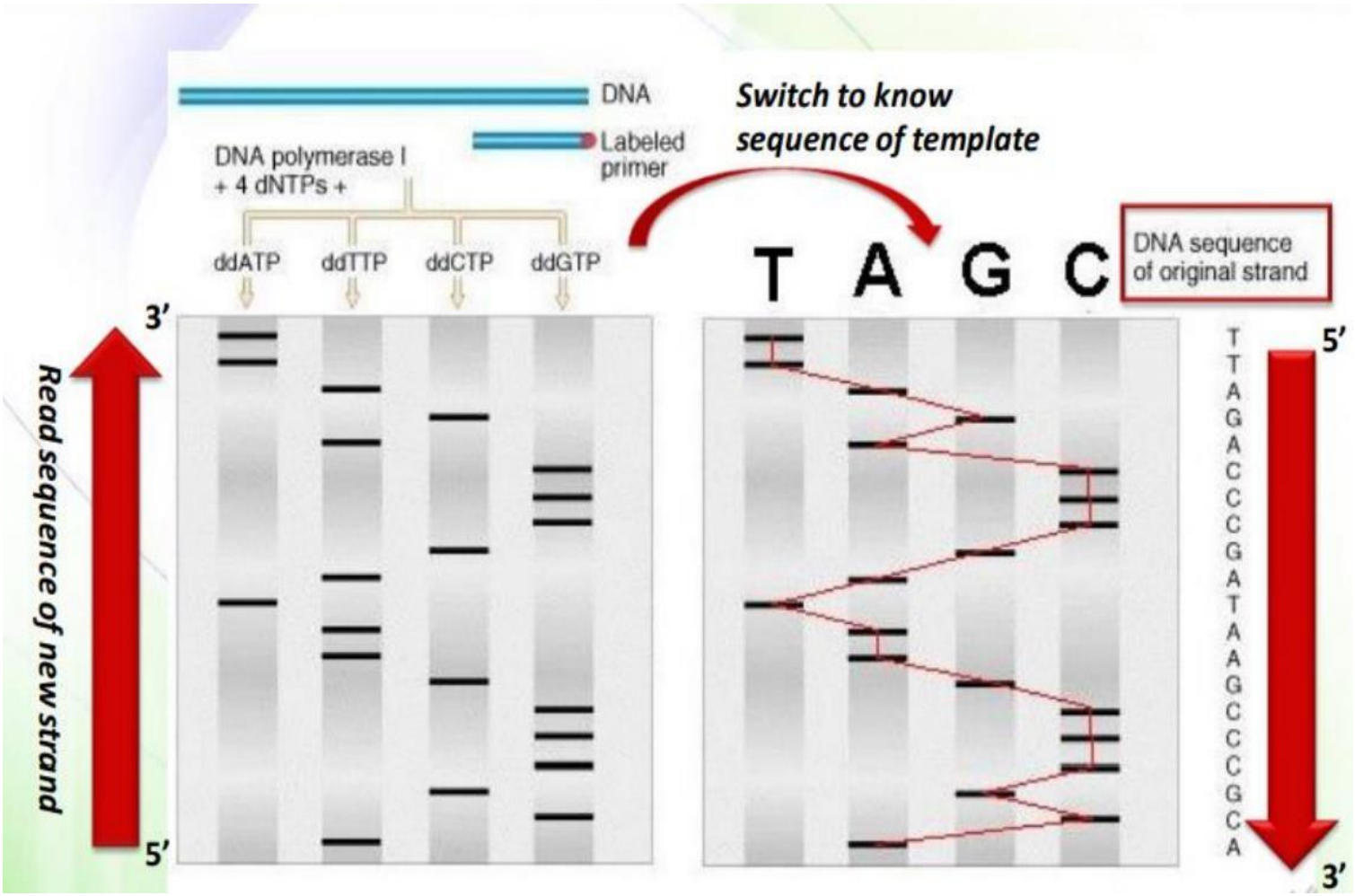
The large DNA in the top , the small in bottom and these differ by just a nucleotideand these resolutions of gel are high so we can separate DNA fragment based on size even if they are different by one nucleotide.

By going from bottom to top we would beable to know the sequence of the newly synthesized DNA so we can know the sequence of template.

Note: that above we started at the 5 primeend because whenever the A is added it's closer to the 5 prime end.

So the sequence of the newly synthesized DNA would be (ATGTCAGTCCAG) and that is5 prime end to 3 prime end , and that's howwe know the sequence of the DNA.

Note: the length of different DNA fragmentsis inversely proportional with respect to theirrate of movement.

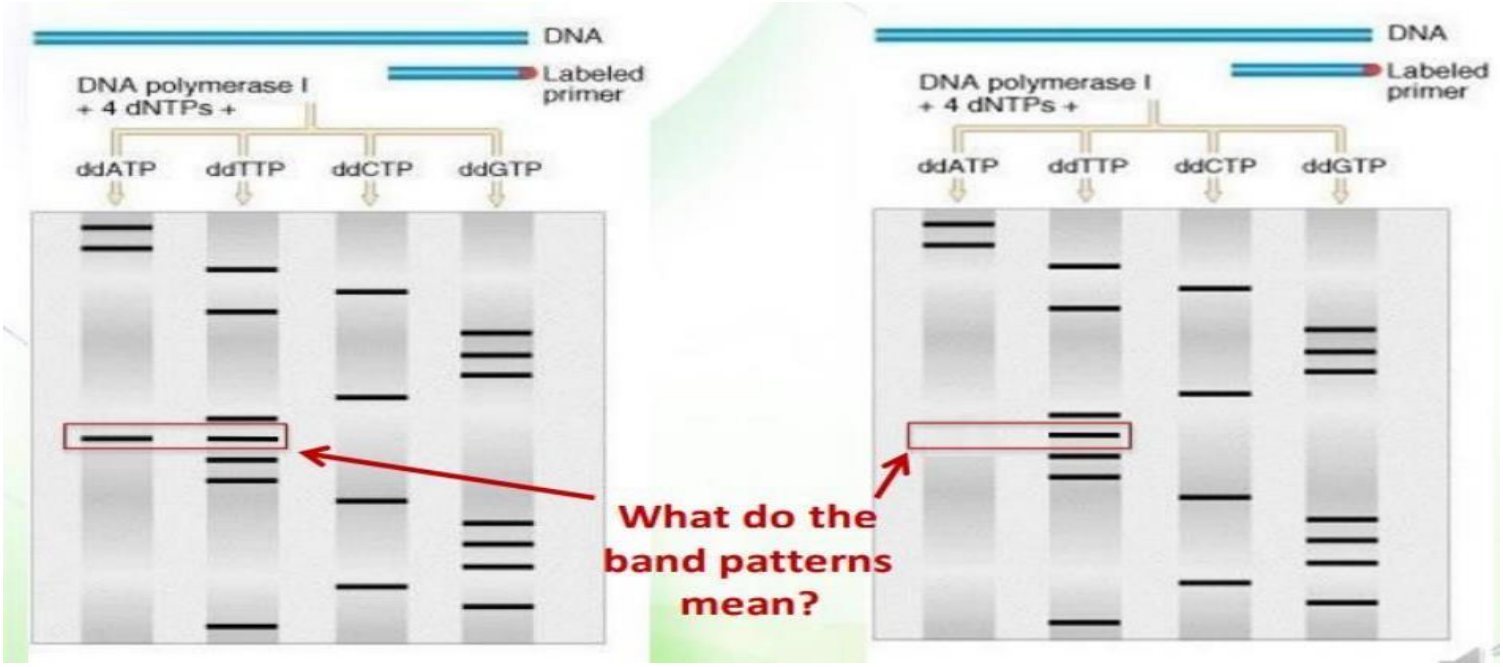


-Showing above the gel done synthesized DNA .

- on the left side of the picture the gel represents the fragments of the newly synthesized DNA and their sequence which is TGCGGGCTTATCGGGTCTAA.

Moving from the 5′ end to the 3′ end.

- Hypothetically, by switching the gel (sort of like) we would be able to tell the sequence of the template by switching the letters instead of A we would have T and so on starting reading from the top.



In the gel on the left there are two bands at the same level, this means that they have the same exact size, this is possible because human cells are diploid, , that is they have two copies of each chromosome: one from the mother(maternal)one from the father(paternal). meaning that there is possibly a variation of one nucleotide between the mother's DNA and the father's DNA (Heterozygosity) which happens to be the DNA sequenced in the previous image. So DNA polymerase is reading the same nucleotides from both DNA from the individual, but when there's a variation in one nucleotide, the mother's DNA is read (T) and the father's is read (A) for example, so both DNAs are read and both fragments would be detected on the gel at the same site because they have the same exact sequence differing only in one nucleotide also known as Polymorphism or SNP (single nucleotide Polymorphism).

-So you can have DNA fragments of the same length migrating together but in different tubes and it can be considered as Polymorphism or a mutation in one chromosome.

-How do we differentiate between mutation and Polymorphism?

If it exists in more than 1% of the population it is a Polymorphism , if less the it's a mutation.

Explanation of the observation on the Right:

• It could be a mutation on both chromosomes (mother's and father's) or maybe a variation in both chromosomes can cause changing in the nucleotides to ddTTP rather than ddATP for example (Homozygosity)

-Then we have to study if this genetic variant or mutation is pathogenic or not ,but in the case of

the individual in the left he's (Heterozygous) or a carrier depending on the change whether a mutation or polymorphism.

Fluorescence-based DNA sequencing

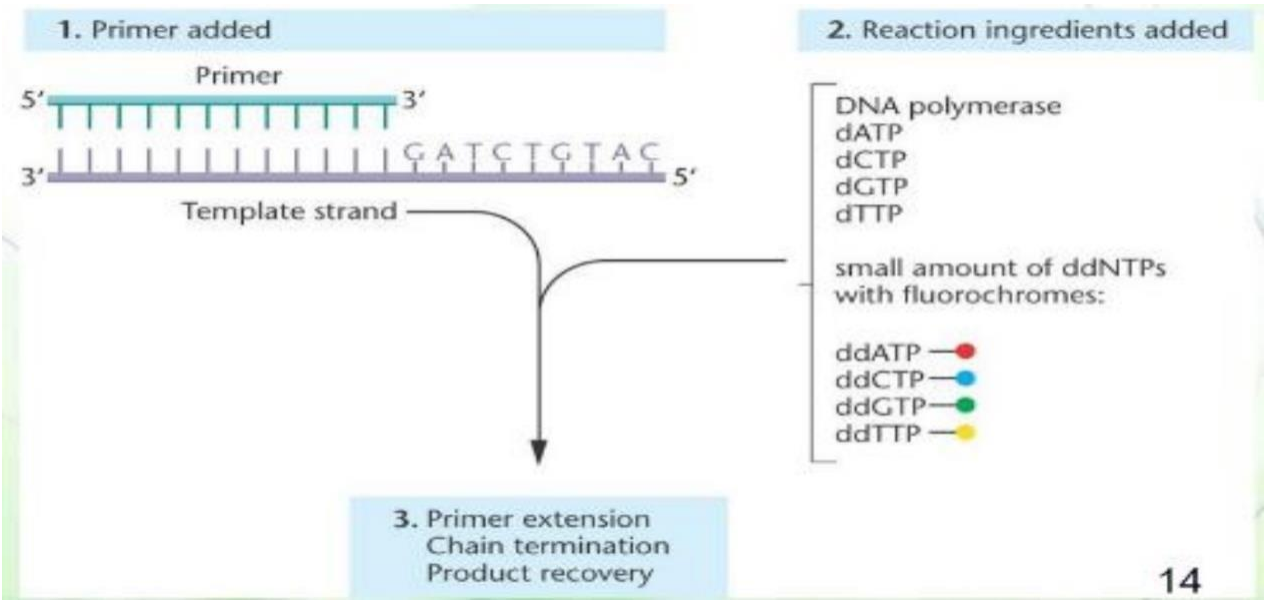
-Firstly, working with radioactivity is not really friendly (It's hazardous), as the radioactive phosphorus might enter his / her body inducing mutations (overall it's harmful).

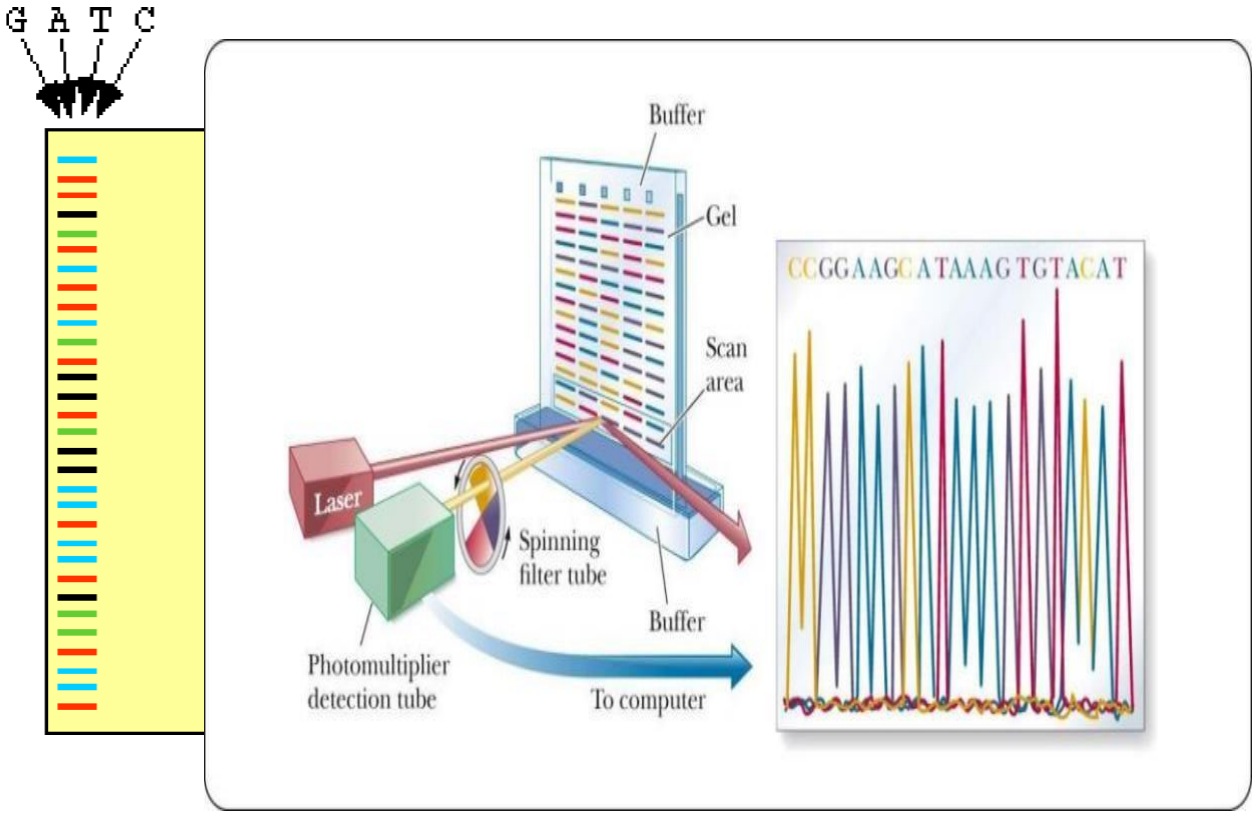
-So instead scientists decided to use something less harmful like fluorescence + automated (letting the computer read it or any other instrument) which is-sensitive, less dangerous and less laborious.

-Reactions include the four deoxynucleotides plus the four dideoxynucleotides in the same reaction with each ddNTP labeled with a unique fluorescent tag.

-So in a single tube We're going to have a template, the primer not labeled, the substrates (deoxynucleotides), DNA polymerase and of course the fluorescent substrates (giving different colors) meaning the ddNTPs.

Note : the idea is that when adding dideoxynucleotides synthesis would stop or be stuck while it continues for the majority as in the addition of deoxynucleotides.

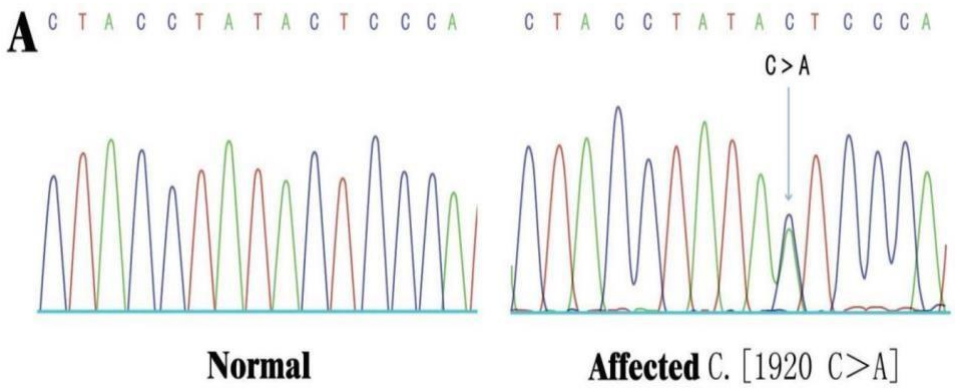




-It's the same exact process, except that we use a single lane not 4 lanes on gel electrophoresis since every type of the 4 types of ddNTPs emits a unique colored signal, and a detector would read and analyze the signal that would come out of each band, thereby identifying the DNA sequence since every florescent signal represents one type of a ddNTP.

- The computer software would give us the signals in the form of peaks as shown in the picture, for example: the two yellow signals are read which means that the computer read C C dideoxynucleotides . So now we have an instrument that reads the DNA sequence by translating colors into letters meaning that each letter has its own color represented in such a peak.

But **what if there's a variation or a mutation? how would it look like?**



What does it mean?

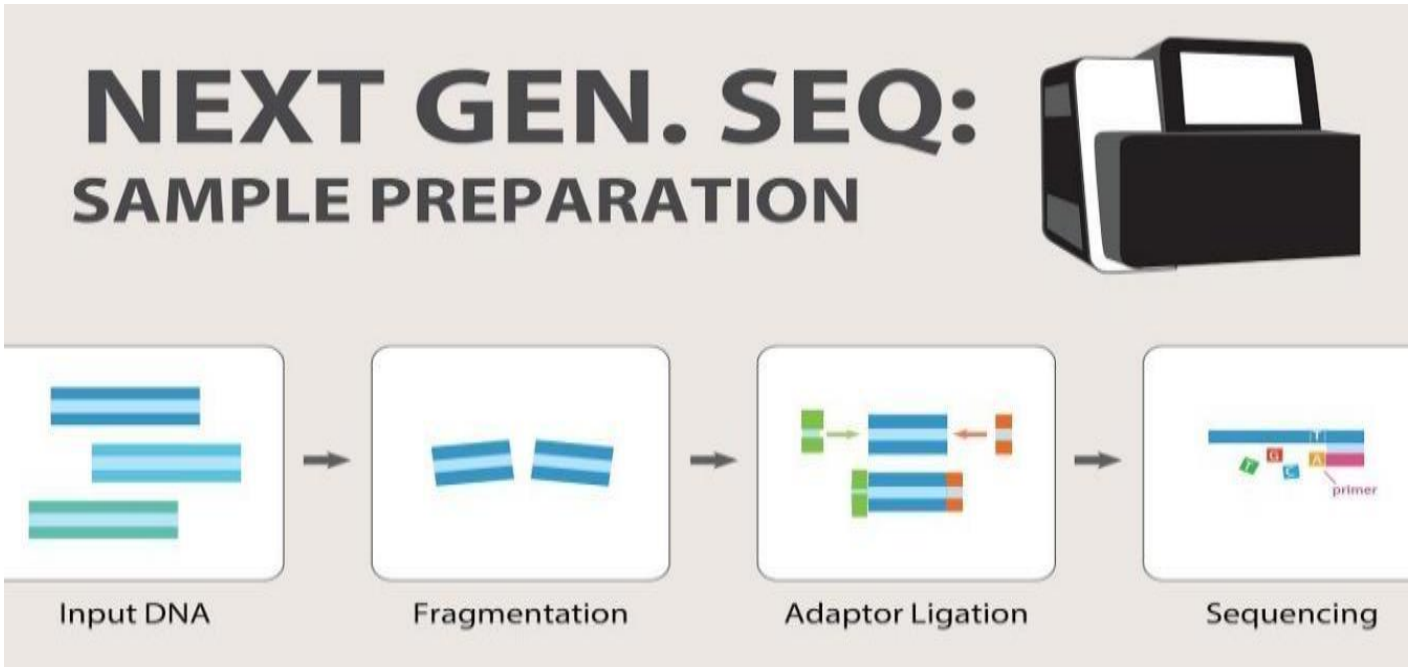
- Normally, we have peaks of different colors, but when two peaks overlap, the computer detects peaks of similar height, and interpret it as polymorphism which means there is a variation between the mother's and father's DNA just like we mentioned before. The signals would come from both the mother's and father's DNA, if the mother's DNA is read C and the father's G at a certain position (Heterozygosity), the DNA polymerase would read and add both nucleotides to the newly synthesized DNA fragments, some fragments would have G nucleotide, and others would have C nucleotide, but both portions are equally the same in terms of their lengths which explains why the computer would read both dideoxynucleotides the same in the form of two peaks.

What if there is a mutation on both alleles , for example instead of having a G on each allele, a C would be present instead ?

- we would still have a single peak but representing a different signal than what would normally exist .

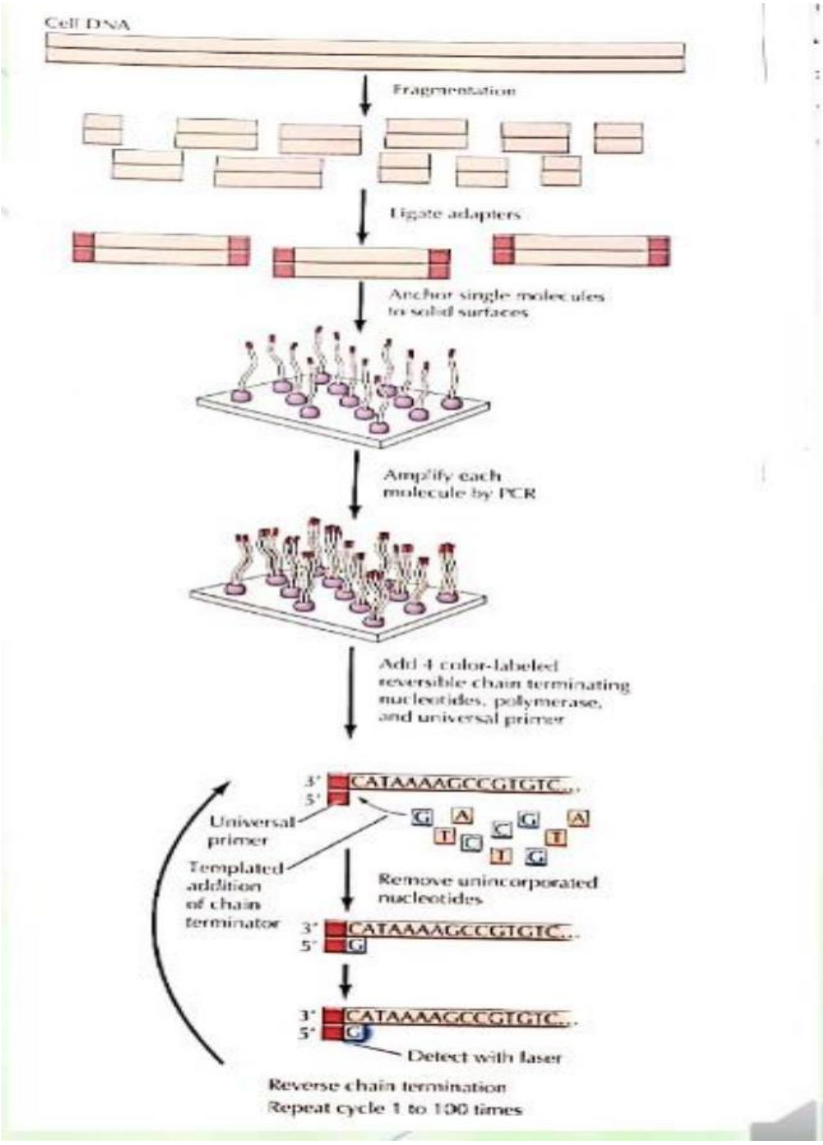
Note: If the cold color changes we expect it to be a (T)-red, but then it turns out to be blue that is a (c), meaning that we have a homozygous individual, having a (c) on both chromosomes instead of having a (T) on both chromosomes

What is Next-generation sequencing?



Next generation sequencing

A group of scientists revolutionized DNA-sequencing by developing this technique in 2007 , including the scientist Greq-venter.



The principle of Next-generation sequencing: -

- Cellular DNA is fragmented randomly (notice that fragments are of different sizes)
- Identical DNA adapters are added enzymatically at the ends of each DNA fragment, in order to :

1- Allow DNA fragments binding a special platform (a solid surface)

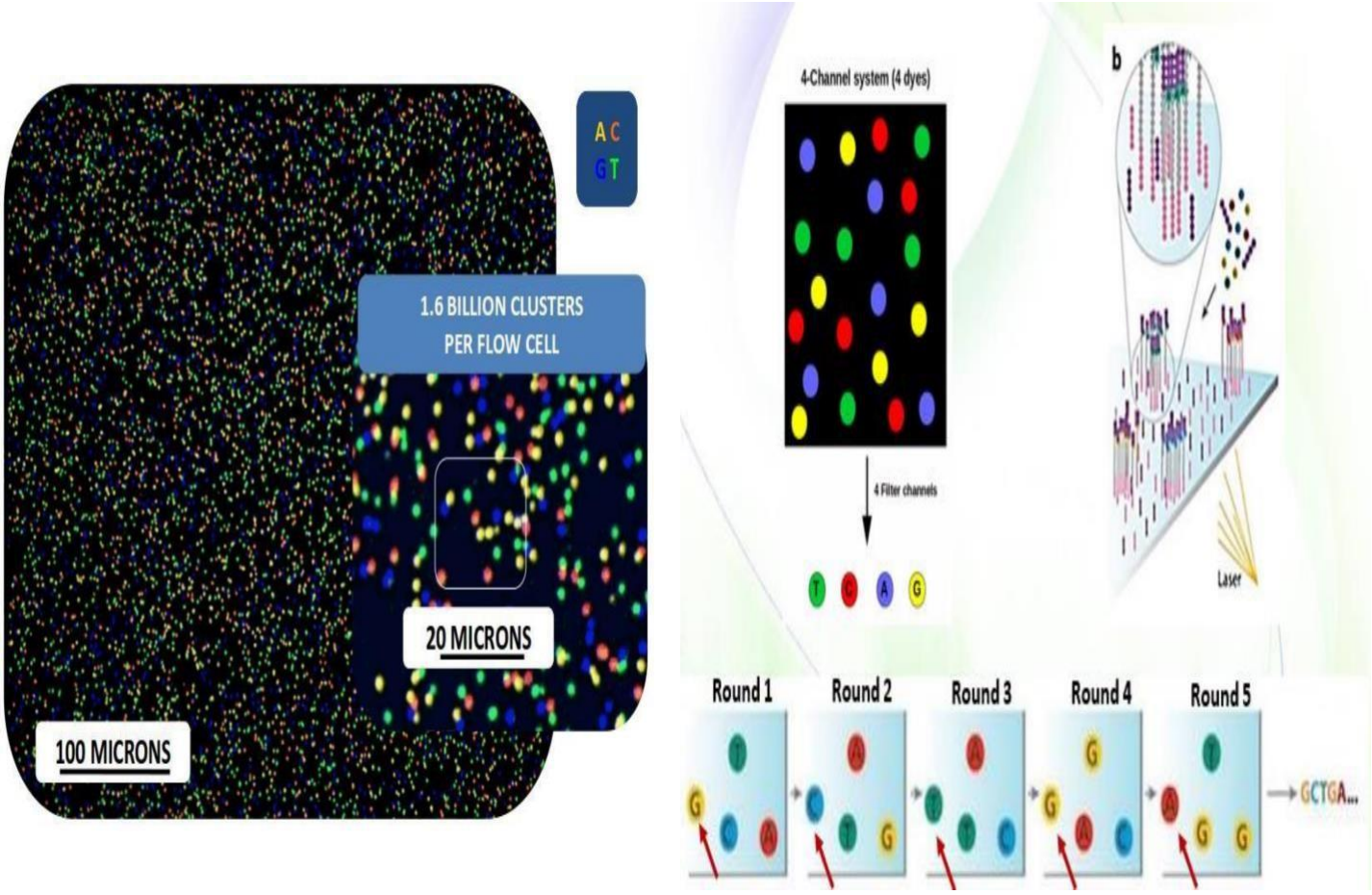
2- They can act as binding sites for primers

3- DNA adapters can be used as tags to differentiate in case we are sequencing 2 different samples.

- Each DNA fragment is attached to a solid surface and amplified (multiplication of each fragment) like PCR Using primers that anneal to the adapter sequences (Which are similar for all fragments)
- Four-colored nucleotides with terminating ends are added .

Note: that these nucleotides are special nucleotides NOT deoxyribonucleotides that have the ability to fluoresce in 4 different colors and the added substrate should be activated before the addition of a new nucleotide, then this nucleotide is activated (modified) by light rays : (1) releasing a fluorescent light (signal) which will be detected by a special camera (2) allowing the addition of a new nucleotide.

- The cycle is repeated.



- We perform the previous steps in multiple rounds and in each round, a special camera-system records a fluorescent signal coming from the nucleotide until we sequence the entire fragments. (The entire human genome could be sequenced in this method within 24 hours).

Notes :1-The signals are transmitted and detected within a small period of time. Furthermore , the florescent signal can only be transmitted provided that the Nucleotide is activated (It is excited by a light source).

2-The addition of the subsequent nucleotide is very rapid because it is an enzymatically activated reaction and these reactions are known to be fast.

In the left Image, the camera system detects all of these signals in the same time and after that it organizes these signals in the form of a DNA-sequence.

Final-Note : the following video may contain extra details that are not required.
https://www.youtube.com/watch?v=womKfikWlxM

The End