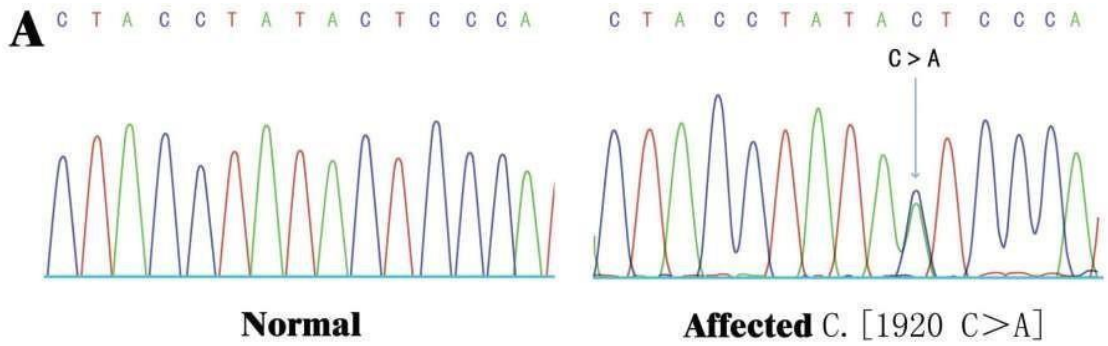
Genetics & molecular biology

◉Sheet ○Slide

Number:	#6
Done by:	رشا عطية
Corrected by:	-Abdullah alaraj
Doctor:	-mamoun ahram

This lecture is written from section 1, at the beginning of the lecture the doctor gave us a chance to ask him any question about the previous one, and here is the questions:

1.



In the affected plot we have two curves that are crossing each other, this indicates that we may have a mutation or a normal variation in that particular region of the genome, **the question was how we can know if It is a mutation or not?**

This depends on experiments and researches, so they have done the sequencing for the human genome by choosing 5 people. **Do the Sequences of their genome normal or not? How would I know?**

A) This is done by sequencing as many DNA molecules as possible.

For example:

we do globin gene sequencing for a group of people , many of them have a sickle cell anemia and the others are healthy individuals , we can see that all healthy individuals have (A) nucleotide in specific position in globin gene , while who has a sickle cell anemia has (G) nucleotide ,therefore we indicate that the difference in that position
May greatly be a mutation that leads to sickle cell anemia

B) is that difference in the sequence will change the structure of the protein or not (by changing the amino acid)?

We will translate the DNA codes into amino acids, so if the amino acid changes from glutamic acid into valine, this gives me a sign that this is a mutation not normal variation.

C) The functional experiments (the strongest sign)

For example:

Some people can't take aspirin because this will lead to bleeding, I say that this may have happened because of a mutation in a specific gene that codes for an enzyme which is responsible for the metabolism of aspirin (this is a hypothesis).

After many studies I found that all people who developed bleeding when taking aspirin have a variation in a specific position , **is that mutation or Variation?**

This is **not enough**, we make functional assays, and I take the cells and change the gene from normal to abnormal, or from abnormal to normal "This is a mutation studies", the question is … **does the changes that I make cause lack of metabolism for the aspirin or no?**

If **yes**, it's a mutation.

Important notes:

1. In the sequencing and PCR, the reaction happens with big number of templates, polymerases (they work all at the same time) and primases.

* We don't talk about one molecule, we talk about many molecules.

2. Can we do one copy of the DNA?

Definitely we can by **single cell analysis** (it means that they took just one cell and extract its DNA to study it), but actually when we want to study DNA we take big number of cells and extract all the DNA they have.

(In dinosaur we take small # of DNA).

3. The primer will clearly identify the region that I want to amplify, both of primers should be hybridized to the DNA surrounding that specific Region.

4. The higher temperature the higher accuracy

(More accuracy means that the hybridization between the primer and the DNA is specific), in low temperature we can have imperfect hybridization, but if we increase the temperature the H bonds will be Broken due to heat.

*If we increase the temperature too much the amplification won't happen due to the breaking of H bonds even in perfect hybridization.

*To know the perfect temperature that helps me to get the specific and unique band, we use PCR and do different ran in different temperatures.

* the difference in one degree or even half a degree may result in a huge difference

For example:

The range of annealing temperature is from 50 to 60 degree, so I start doing different ran in 50, 51, 52 … to 60, then I see which one of these ten ran gives me the best specificity (gives me a single band that must be the predicted one, I know that my product length must be 1 kilo base), so if the products length is 2 kb that means that the amplification is specific but the product is not what I want.

*this helps me to know that there is a mistake which is using the wrong primer.

* This gives me an idea about the other nonspecific region that the primer amplified has a sequence which is **similar** to the region that I want

*this may happen to genes from the same family.

For example:

Histones 1 and 2 or 2a and 2b, the sequence for 2a and 2b DNA is highly similar.

The last question:

How can I know the specific sequence of the primer that I have to use (the dry said that this is an extra in)?

We have a database for the human DNA sequence, so I take the sequence of the gene that I want to amplify, then I choose the primer depending on

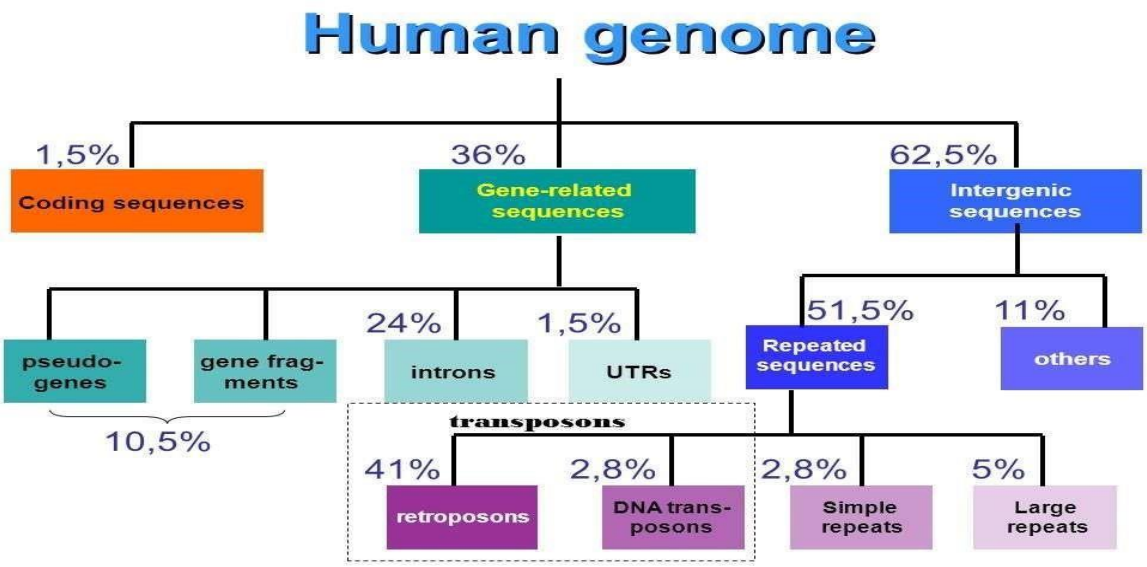
1- G_C content

2- the length of the product

3- the region

Then I take these primer and check that their temperature is suitable for functioning affectively , after that I do a search for the database (are the 2 primers going to amplify for another region or not?) if yes so they are bad primers .

I tried my best to simplify the answers as much as I can ^_^



*we have 3 billion base pairs and when we look on the human genome we found that:

1- 1.5% of the whole genome is a coding sequences (which is nth compared to others component of the genome).

*coding sequence: it is the sequencing that are transcribed (making RNA) to be translated to proteins.

*The scientist in the past called the remnant of the genome which percentage around 98.5% **a junk DNA.**

2- gene-related sequences : sty related to genes but do NOT Coded for a protein such as introns, untranslated regions and Pseudogenes.

* Pseudogenes: they are look like genes but they are not transcribed (through evolution they lost their ability to be Transcribed).

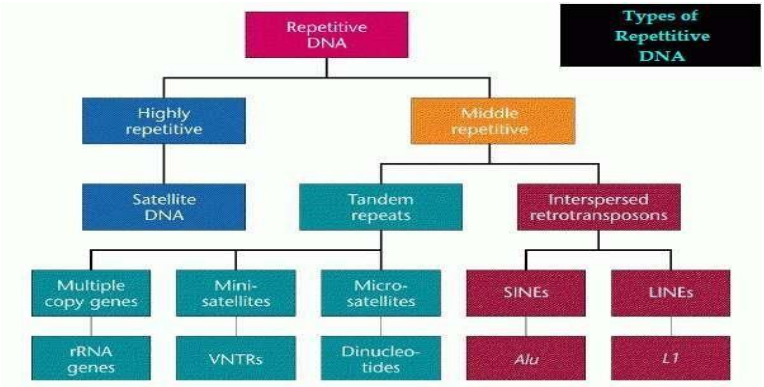
3- Most of our genome is basically intergenic sequences.

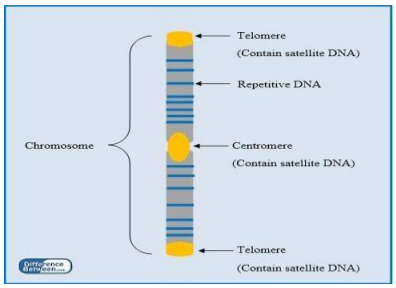
*Refer to the mind map above to know its divisions.

A_ repetitive sequences

B_ others

Repetitive DNA sequences

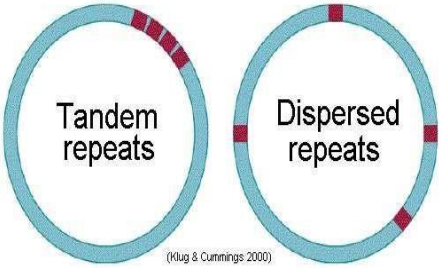




The repetitive DNA region could be:

1-highly repetitive (they call it satellite DNA): we can find it in **telomeres** and **centromeres**.

*notice in this pic where the telomere and centromere is located



2- Middle repetitive (the repetition is lesser than highly repetitive):

*They can be different repeats:

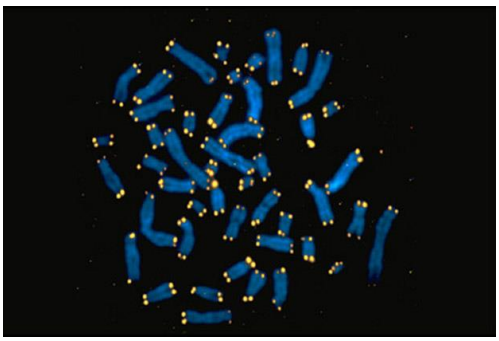
A_ tandem repeats

B_ interspersed retrotransposons

Now we will discuss the repetitive DNA with details:

***highly repetitive**: **satellite** (**macro-satellite DNA**): they are Regions of 5-300 bp and they are repeated 10^6-10^7 times such as telomere and centromere. (They make up 10% of the genome).

* Centromere repeats (171 bp) are unique for each chromosome so they used it a technique called **fluorescent in situ hybridization** in order to color the chromosomes by using probes specific to each centromere so we can identify the chromosome, for an example we see 2 centromeres that means that the chromosome is **intact and okay**, we can count the labels as well.



*In centromere these repetition is different from

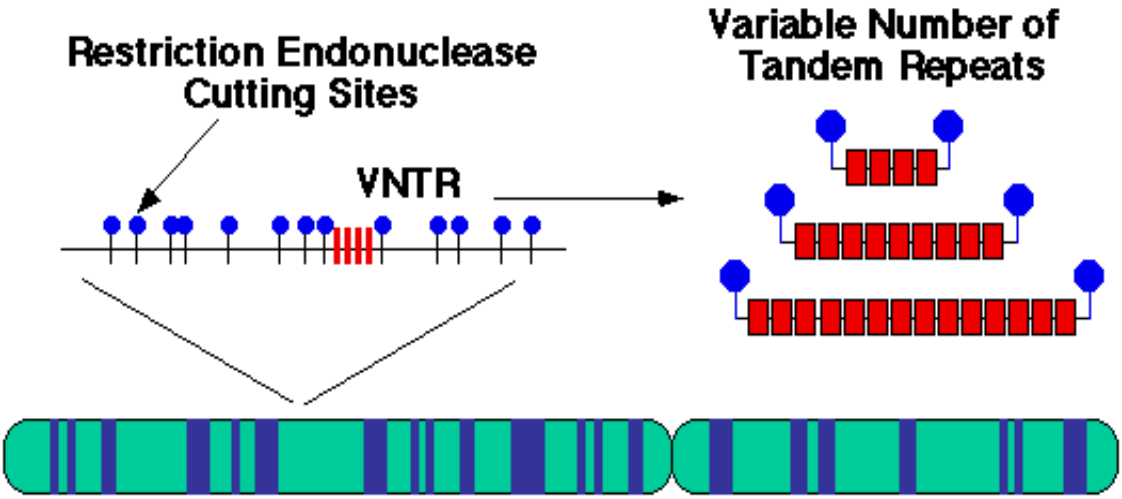
One chromosome to another, therefore we can Label the whole chromosome based on Centromere (because they are specific).

***middle repetitive:**

A_ tandem repeats:

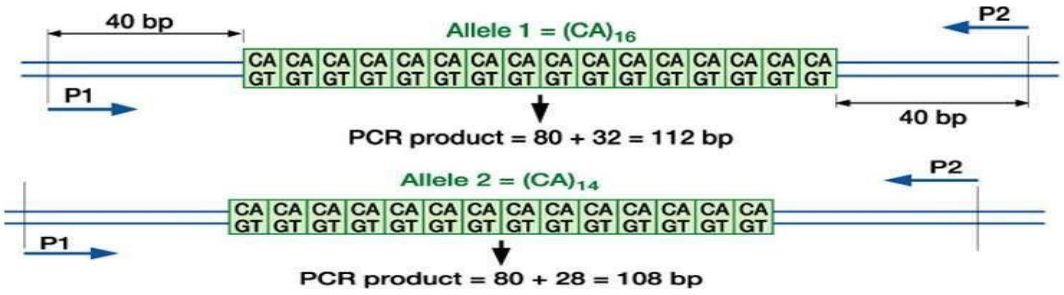
1. (**Mini-satellite DNA**): an example of them sty called VNTRs (variable number of tandem repeats), they differ among individuals by the Number of repeats.

* Two persons MAY have the same number of repeats.



2. **micro-satellite DNA:** they are smaller than the mini, an example of them is STRs (short tandem repeats) which its length is from 2 to 10 bp repeated 10100 times.

*For different individuals, it is located in the same region of specific chromosome but with different number of repetition.



This pic is an example for the previous point, (allele 1) has (CA) which is repeated for 16 times, while the other person (allele 2) has (CA) which is repeated for 14 times.

---

1. Polymorphisms of VNTR and STR:

We have differences with each other, **so for example:**

Person has 10 repeats at specific location in his chromosome while Other one has 20 repeats.

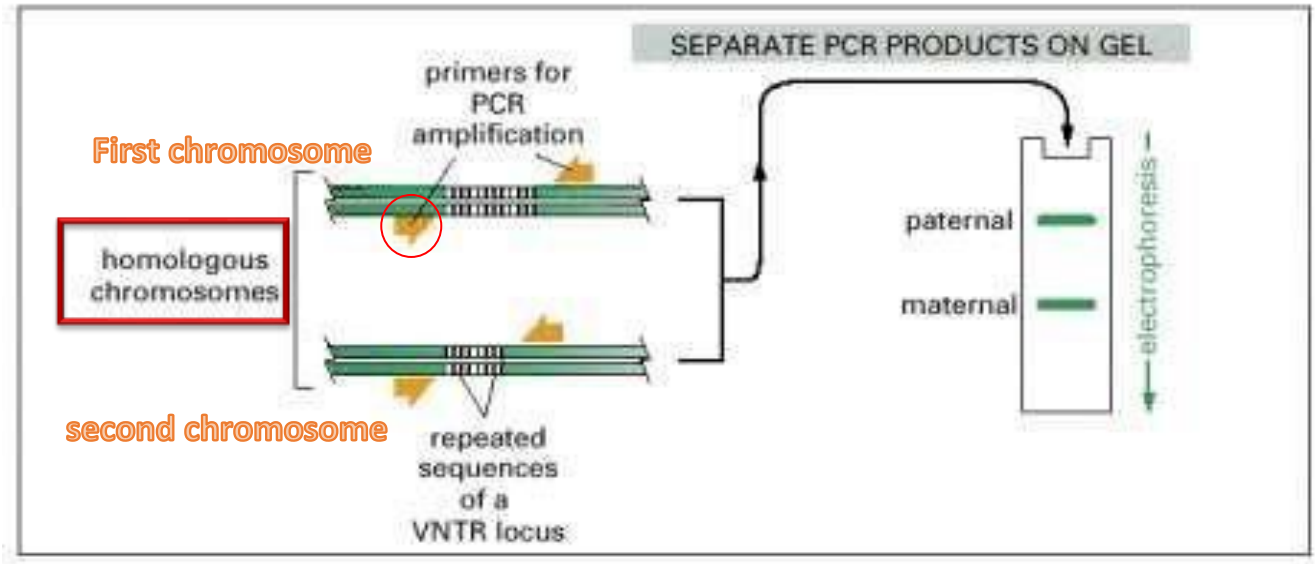
***Remember** that we are diploid because we have two chromosomes , one of them is maternal and the other is paternal , in this specific chromosome whatever its number is **for example** chromosome number 5 we have STR or VNTR , but the paternal chromosome has 10 repeats while the maternal chromosome has 5 repeats , so we do PCR for the region that contains the repeat (the whole repetition region in each chromosome) and we use the same primer because they are at the same # of chromosome in specific position so the sequence is similar.

*the primers are found outside the repeat region, the primer amplified that region but the # of repeats are different in each chromosome , therefore the product that will be produced from the first chromosome is larger than the second one … **why**?

Because they are different in the # of the repeats **but** they aren't differ in their sequence.

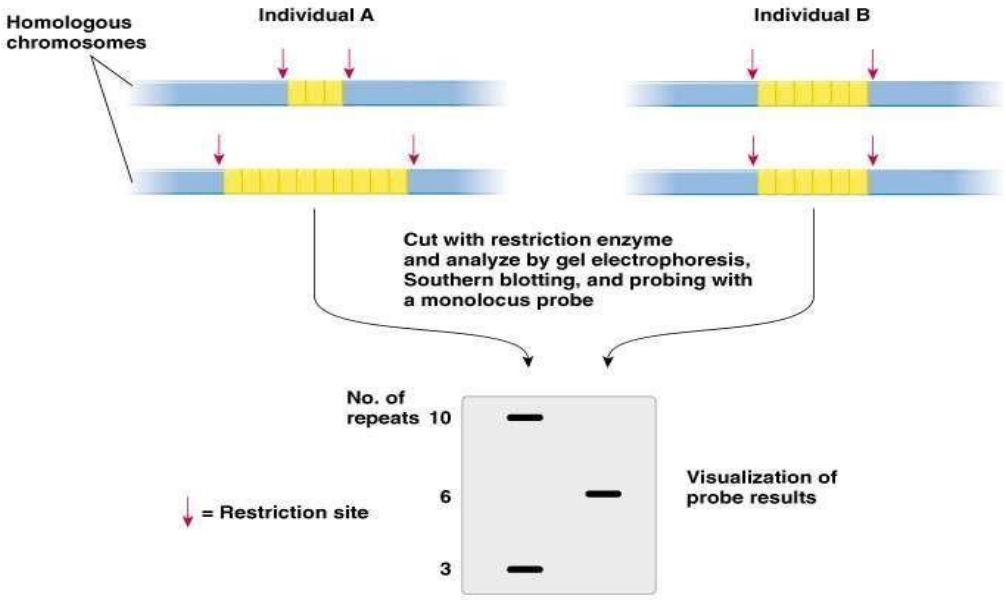
*after doing the PCR we do gel electrophoresis that will show me two bands.

Please focus in this pic for further understanding:



Notice the location of the primer (it is not attached on the repeated region at the beginning **,its location is outside**)

Here in this pic we do PCR for the same region of VNTR or STR of two people (the explanation is below the pic so please don't skip it):



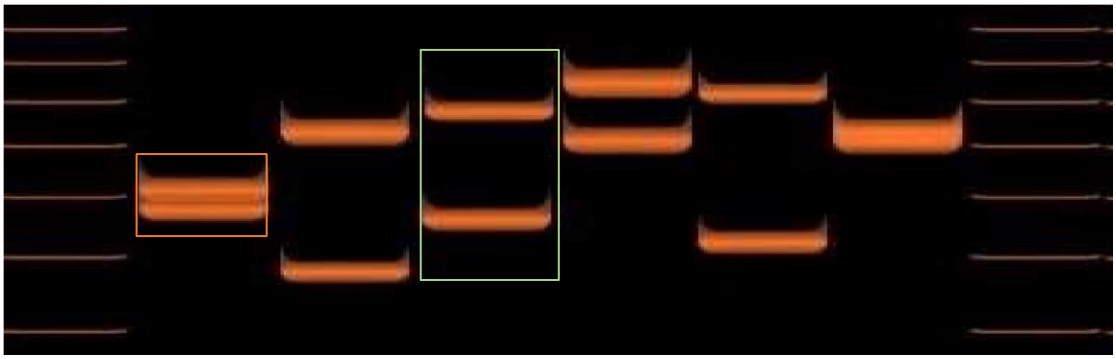
Individual A:

The first chromosome's number of repeats is 3 while the other chromosome has 10, so when I do amplification (PCR) for them, I will Get 2 bands.

Individual B:

Both chromosomes have the same number of repeats (the same length), so this will give me one product (one band).

The picture below illustrates VNTR allelic length variation among 6 individuals.



*This pic shows us that each person has different repeat from one another.

*in the same person, both chromosomes could be similar to each other but they are different, like in the orange box above.

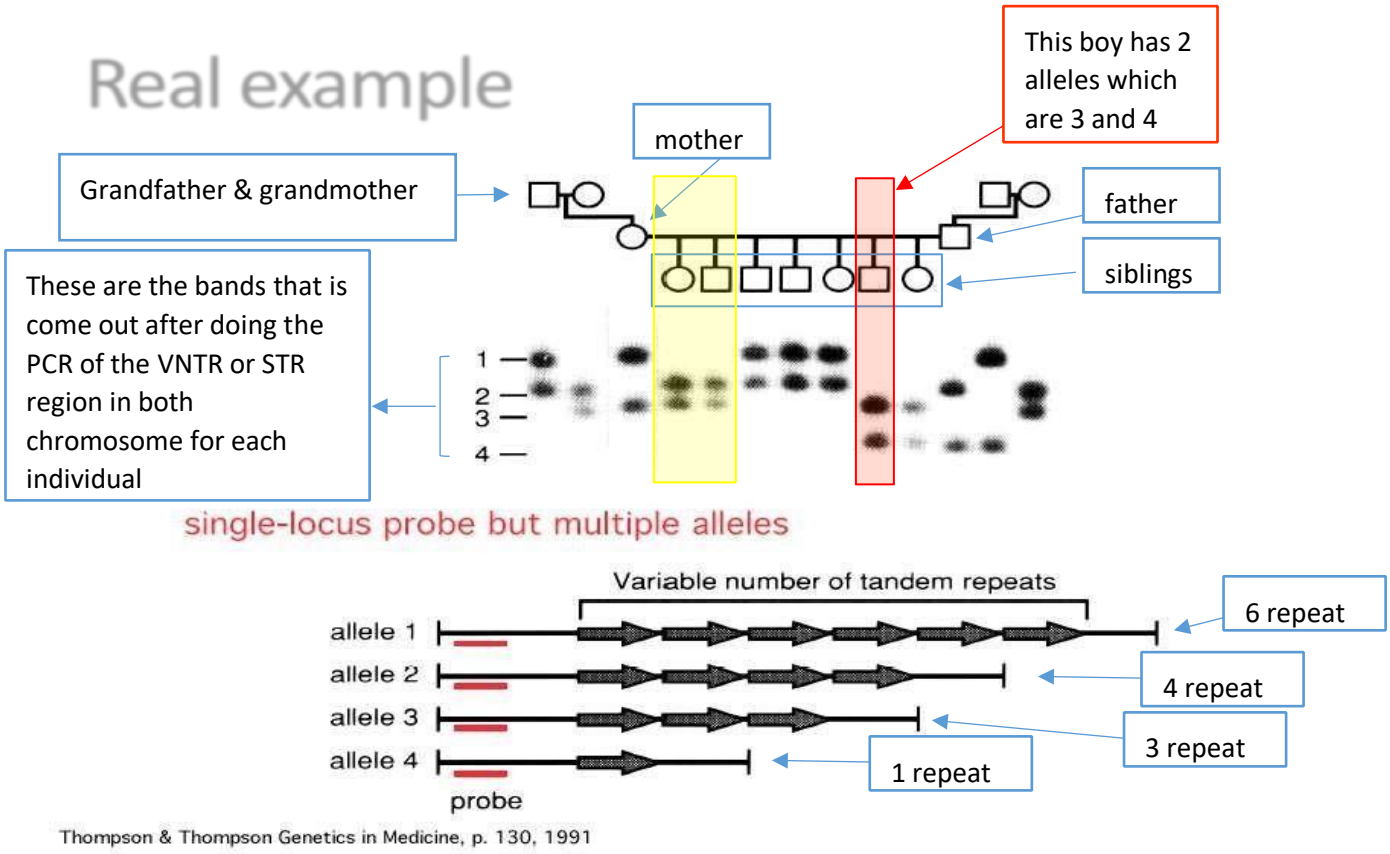
And they could be really variable like in the green box.

***The likelihood of 2 unrelated individuals having same allelic pattern** extremely **improbable**

---

*We can differentiate between people by the differences of the number of repeats, and here is a **real example**:

***But before further explanation, I added most of the information in the pic itself so read them carefully.**



*The boy above in the red box has 2 alleles, allele #3 is taken from his mother (his mother took this allele from her mother), while allele #4 is taken from his father (the father took this allele from his father)… and So on (the same concept for each one).

*The yellow box indicates two siblings one is a boy while the other is a girl, both of them has the same alleles (the same pattern).

*Suppose that one of them kill someone (we have a sample of blood that has been taken from the place of the accident) ; hence we check the PCR and we saw that both of them has the same alleles like the example above , what could we do to know the killer (and differentiate Between this two)?

1. take DNA sample from both of them

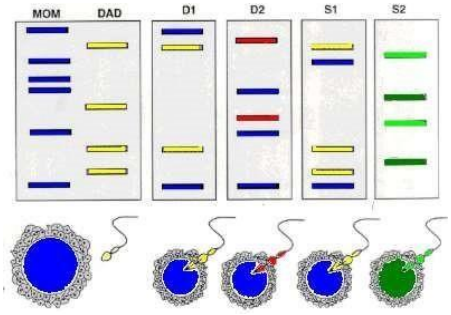
2. do PCR for another STR or VNTR in other regions

*they must have different alleles for different STR and VNTR because we have thousands of them (we can find them in other regions and other chromosomes)

*we can know the killer simply by detect the gender

*if they were identical twins!!! (The dry said we can know by fingerprint)

*we can use this methods for paternity testing **(we don't take only one region , we take many regions to have more accurate result).**



---

2. Single nucleotide polymorphism (SNPs):

*another source of genetic variation.

*a change in one single nucleotide in a certain sequence.

* SNPs occur throughout the human genome - about one in every 300 nucleotide base pairs.

***what is the difference between mutation and variation, how can I differentiate between them?**

A) Mutation (SNP) happens when changing a specific nucleotide is associated with certain **disease**.

* If it is not associated with a disease we call it variation **(it can be sometimes harmful when the change affected your metabolism of a certain drug.)**

B) To consider sty a (SNP), it must be found in the population with more than 1%

*If the percent is less than 1, we call it a mutation.

*If the percent is more than 1 percent, we call it variation.

*They are not random they are in specific positions.

*in our genome we have 10 million SNPs, which means 10 million variations BUT THE QUESTION IS, **does all of them are important?**

NO, the number of the significant and important SNPs is 0.5 million only

Categories of SNPs

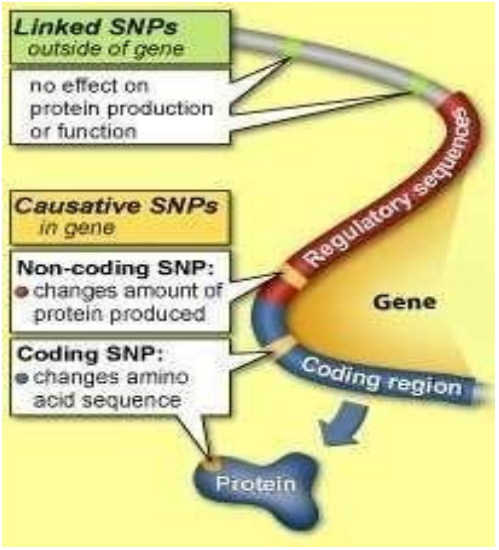
***what are the important SNPs?**

A) Causative: a variation causes changes in a phenotype (could be found in the coding regions or non-coding regions)

For example:

1. 25% of people who take aspirin they have bleeding (when we look at their DNA sequence, we will find many polymorphisms that cause a phenotype).
2. Some people who drink coffee may become worried while the others drink it and go to sleep, **why this happens even they have the same metabolism enzyme?**

This is because there is a polymorphism that make people more sensitive than others.
*these polymorphisms can increase or decrease the sufficiency of certain enzyme.



*the causative SNPs could be found in the coding regions or non-coding regions:

1. **In coding region**: they change a nucleotide, therefore changing an amino acid in the enzyme.

For example:

I have leucine and isoleucine, both of them is branched, non-polar and aliphatic .As well I have two enzymes, the first one has a leucine amino acid at specific position while the other has isoleucine, **does the enzyme activity change?** Maybe but not so Much.

2. **In non-coding region:** such as in the promoter or the enhancer or intron (there won't be any changes in the amino acid of the Protein, but there is a change in the phenotype), it will affect how much protein we have in the cell.

B) Linked: they are linked to a phenotype but they don't by themselves cause a phenotype.

For example:

We take a specific individuals from the population, 10% of them have high susceptibility to produce hypertension, so we asked ourselves **WHAT IS COMMON AMONG THESE INDIVIDUALS?**

We didn't find any specific gene that is directly associated but we find that 70 % of people who developed hypertension have in certain location in their DNA for example (A) nucleotide while Healthy people (30%) they have (G) nucleotide.

*We took different samples of DNA for both (healthy people and people who developed hypertension) and change the (A) to (G) and vice versa, the result is that all of them have the **same phenotype** (this polymorphism **doesn't** change the phenotype but it is **linked** to it).

*So if someone comes to me and in his DNA sequence I find (A) nucleotide , I will tell him that this polymorphism is linked with people who developed hypertension so be careful , you may develop hypertension.

* That doesn't mean he will certainly develop the disease but there is risk so be careful (we recommend that he should check himself every now and then).

*this happens (linked SNPs) outside of gene.

TRANSPOSONS (JUMPING GENES):

*They are segments from the DNA that can move from their original position in the genome to a new location.

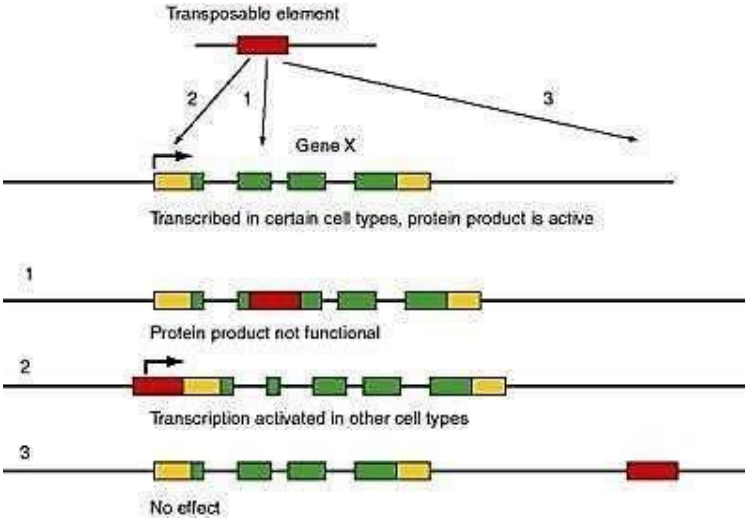
Two classes:

1. DNA transposons (2-3) % of the human genome
2. RNA transposons: or retro transposons (40% of the human genome), they came from retroviruses millions of years ago. (Branch (b) of the middle repetitive page 7)

*Now depending on the sequence they are:

1. long interspersed elements (LINEs)
2. Short interspersed elements (SINEs) E.g.: Alu (300 bp)

*most of the transposons in our genome lost their ability to jump from a location to other.
(They can cause diseases.)



***explanation for the pic here:**

1. If they **enter** the gene they will Disrupt the coding region so The gene will be disrupted.

2. If they came **beside** the Gene **(the regulating region)** It will also disrupt it.

3) If they came to a position **away** from the gene **(no effect)**

*The disease that can occur:
1) Hemophilia A and B
2) Severe combined immune deficiency
3) Porphyria
4) Predisposition to cancer
5) Duchenne muscular dystrophy

***Last note**: rats and pigs transposons have more activity than ours.

Best wishes ☺