بسم الله الرحمـٰن الرحيم

(وَفَوْقَ كُلِّ ذِي عِلْمٍ عَلِيمٌ)

Cytology & Molecular Biology | FINAL 4

The Human genome

Written by : NST

DST

Reviewed by : NST

وَلِلَّهِ الْأَسْمَاءُ الْحُسْنَى فَادْعُوهُ بِهَا

المعنى: الذي لا يموت، وحياته – سبحانه – أكمل الحياة، تستلزم جميع صفات الكمال، وتنفي أضدادها من جميع الوجوه، وكمال حياته يستلزم أن لا تأخذه سِنةٌ ولانومٌ.

الورود: ورد في القرآن (٥) مرات.

الشاهد: ﴿ٱللَّهُ لَآ إِلَٰهَ إِلَّا هُوَ ٱلۡحَيُّ ٱلۡقَيُّومُۚ﴾ [البقرة:٢٥٥].

١٣

اضغط هنا لشرحٍ أكثر تفصيلًا

Quiz for previous lecture

Click here

Molecular Biology (2)

The human genome

Prof. Mamoun Ahram

School of Medicine

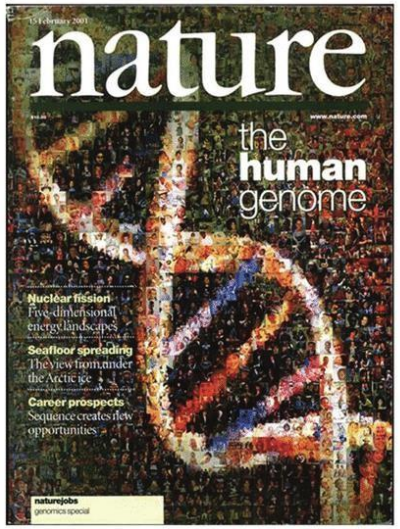
Second year, Second semester, 2024-2025

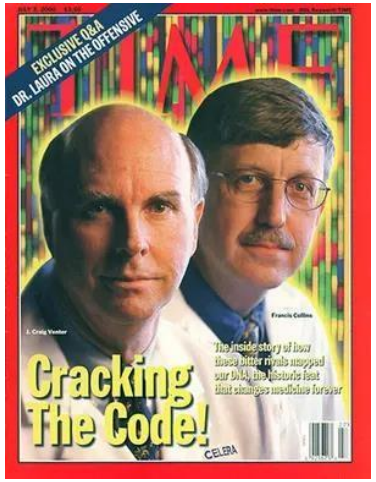
The human genome project

five countries got together led by the United States to sequence the human genome &to know the order of nucleotides in the human DNA in the 24 chromosomes

- A $3 billion, 13-year, multi-national project launched in 1990 led by the US government to (know the) sequence the human genome and to map and identify the genes (a draft was published in 2001 and 92% was completed in 2004).

It was announced by Bill Clinton the president of the United States of America and the project was led mainly by Francis Collins and in the later stages there was competition by Greg venter who established a private company called Celera, they were able to also sequence the Human Genome using the preliminary information generated by Francis.





فك الشيفرة

TO decode the genome in 1990 ,will take from us 15 years,, but Nowadays we have new technology that decode the genome in 4 hours

Major outcomes

- Determination of the number of human genes Also, on which chromosomes are they located, and where are they situated on that chromosome.
- Development of major technologies and bioinformatic tools
- Completed sequences of other genomes
- Open discussion of legal and ethical issues

"If someone decides to decode their own DNA sequence and finds out they have mutations that will lead to a disease, that might sound nice at first. But imagine that this disease isn't treatable – they would spend their whole life waiting for the disease to appear, and you can imagine how much that would affect their psychology.For example, if they apply for a job, the company might tell them, 'Because you have mutations that could cause a disease and cost us money, we won't hire you.'Even health insurance might say, 'You have mutations that could lead to a disease and become expensive, so we won't give you coverage.'And even marriage – they might not get married because it's known that they will eventually develop a disease."

Along with sequencing the human genome, scientists have been able to sequence the genomes of mitochondria and other organisms including advanced living beings like mice and chimpanzees and so on or even simpler ones like viruses & bacteria

Comparative genomics is to look at & compare different genomes of different organisms

Don't memorize the table

organism	genome size (base pairs)	protein coding genes	number of chromosomes
model organisms			
model bacteria *E. coli*	4.6 Mbp	4,300	1 → Circular
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)

Similar number of genes in human and mouse genomes
- we all sort of have the same number of genes overall- we differ in very few genes, so what makes a human human and a mouse mouse?
the noncoding parts regions of the genomes

- Mouse has 40 chromosomes doublets and there are 20,000 protein coding genes; compare this to the human genome where there is 46 chromosomes doublets and there are about 20,000 genes also Some viral genomes are made of DNA, others are made of single stranded RNA and so on, there's a huge variety of genomes.
- When we first started studying in 1990, they said that the bacterium E. coli has about 4,000 genes. So by comparing it to bacteria, people assumed that if E. coli has one chromosome with 4,000 genes, then humans—with 46 chromosomes—should have around 100,000 genes.
- But the more DNA genome sequencing they did, the more they kept reducing the number: first they said 90,000, then 80,000, then 35,000, then 25,000.
- And that's why, when you look at tables on Google, you'll notice different numbers

The ENCODE project (2003-on)

- ENCODE: Encyclopedia of DNA Elements (ENCODE)
- ~75% of the entire human genome is relevant (either transcribed, binds to regulatory proteins, or is associated with some other biochemical activity).

Summary of ENCODE Results	
Protein-coding genes	20,687
Short noncoding RNAs	8801
Long noncoding RNAs	9640
Pseudogenes	11,224
Percentage of genome transcribed into RNA	74.7%
Percentage of genome-binding transcription factors	8.1%

- Near the end of the Human Genome Project in 2003, the US government started another project called The Encode Project.
- They wanted to look at DNA and at the human genome more closely.
- It was found that the percentage of the genome that is transcribed into RNA is 75% instead of the 2% that we were told about previously.
- The protein-coding genes represent about 1-2% of the human genome, but around 75% of the genome can be transcribed into various RNA molecules.
- The mission now is to know what these really do, their function, significance, are they mere noise or do they have a purpose?
- There are other regions that are not transcribed but they are functional and can affect parts of the gene that are transcribed.
- Many of the 75% of the human genome are actually RNA molecules that can be classified into two types: long and short(shorter than 300 nucleotides or longer than 300 nucleotides)

On March 31, 2022…

A gene: a region of DNA that is transcribed.
A transcript: a RNA molecule that is produced by transcription

	Gene annotation
Number of genes	63,494
Protein coding	19,969
Number of exclusive genes	3,604
Protein coding	140
Number of transcripts	233,615
Protein coding	86,245
Number of exclusive transcripts	6,693
Protein coding	2,780

RESEARCH ARTICLE

HUMAN GENOMICS

The complete sequence of a human genome

Since its initial release in 2000, the human reference genome has covered only the euchromatic fraction of the genome, leaving important heterochromatic regions unfinished. Addressing the remaining 8% of the genome, the Telomere-to-Telomere (T2T) Consortium presents a complete 3.055 billion–base pair sequence of a human genome, T2T-CHM13, that includes gapless assemblies for all chromosomes except Y, corrects errors in the prior references, and introduces nearly 200 million base pairs of sequence containing 1956 gene predictions, 99 of which are predicted to be protein coding. The completed regions include all centromeric satellite arrays, recent segmental duplications, and the short arms of all five acrocentric chromosomes, unlocking these complex regions of the genome to variational and functional studies.

In 2022, a paper was published claiming “the complete sequence of the human genome.” But they were lying, because they excluded chromosome Y. The problem with chromosome Y is that it contains many repeated sequences. Imagine an encyclopedia where 50 pages in a row contain only one word repeated: “the, the, the…” If those pages get scrambled, how many copies of “the” are originally there? It becomes nearly impossible to determine the correct order. But a year later, they finally sequenced the Y chromosome and completed the human genome

How many protein-coding genes do we have—the genes that produce proteins? Around twenty thousand, give or take a hundred or so. The mouse also has about twenty thousand. So we’re like the mouse. Drosophila, the fruit fly, has 14,000. Between humans and the fly, the difference is only about 6,000 genes. That’s it. And there’s not much difference between us and a mouse in gene number. So what makes a human a human and a mouse a mouse?

On August 23, 2023. It is finally done.

nature

Article | Published: 23 August 2023

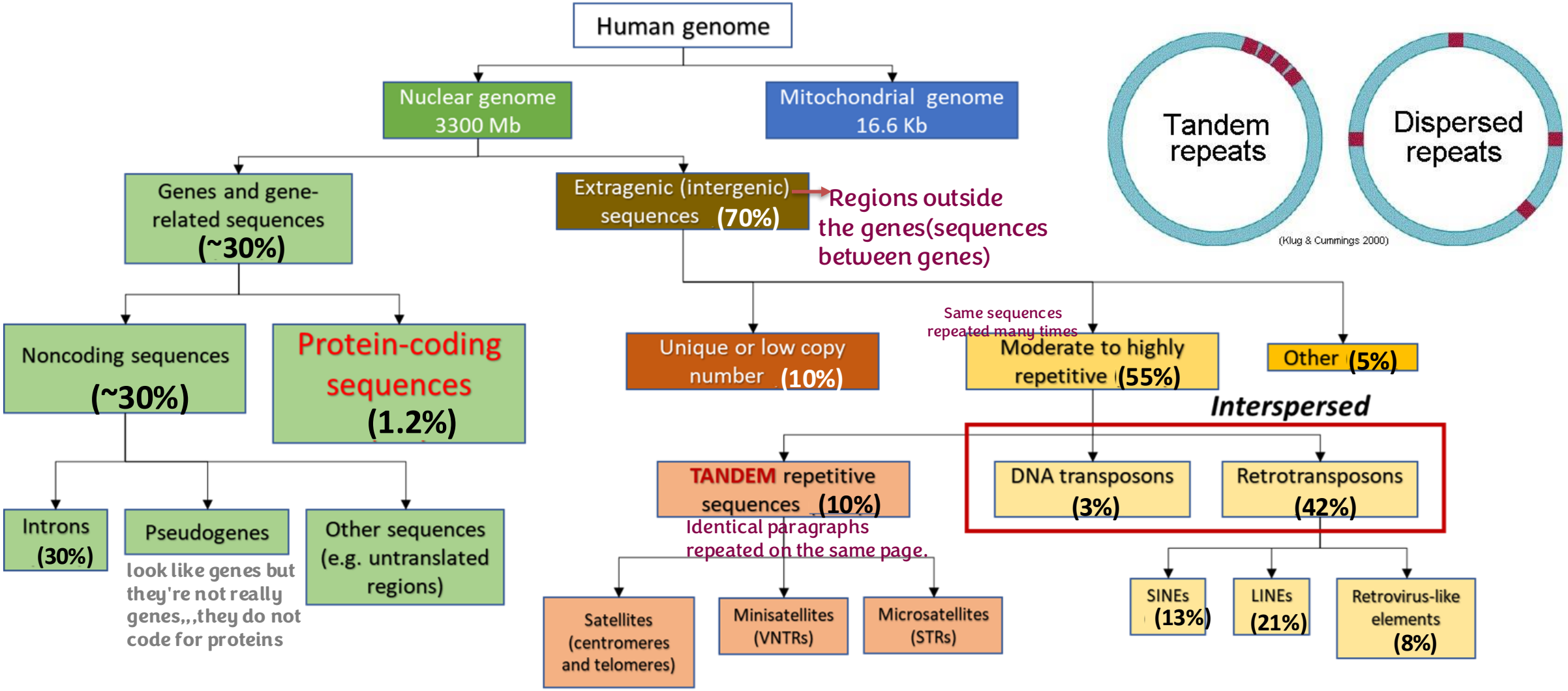
The complete sequence of a human Y chromosome

Arang Rhie, Sergey Nurk, Monika Cechova, Savannah J. Hoyt, Dylan J. Taylor, Nicolas Altemose, Paul W.

Abstract

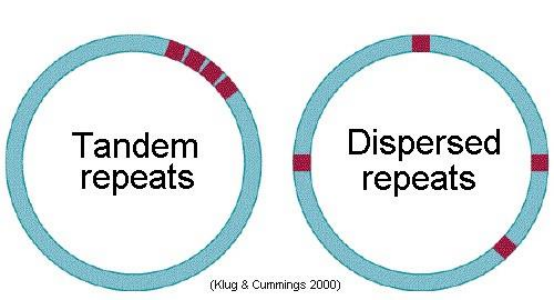
The human Y chromosome has been notoriously difficult to sequence and assemble because of its complex repeat structure that includes long palindromes, tandem repeats and segmental duplications[1,2,3]. As a result, more than half of the Y chromosome is missing from the GRCh38 reference sequence and it remains the last human chromosome to be finished[4,5]. Here, the Telomere-to-Telomere (T2T) consortium presents the complete 62,460,029-base-pair sequence of a human Y chromosome from the HG002 genome (T2T-Y) that corrects multiple errors in GRCh38-Y and adds over 30 million base pairs of sequence to the reference, showing the complete ampliconic structures of gene families *TSPY*, *DAZ* and *RBMY*; 41 additional protein-coding genes, mostly from the *TSPY* family; and an alternating pattern of human satellite 1 and 3 blocks in the heterochromatic Yq12 region. We have combined T2T-Y with a previous assembly of the CHM13 genome[4] and mapped available population variation, clinical variants and functional genomics data to produce a complete and comprehensive reference sequence for all 24 human chromosomes.

Components of the human genome

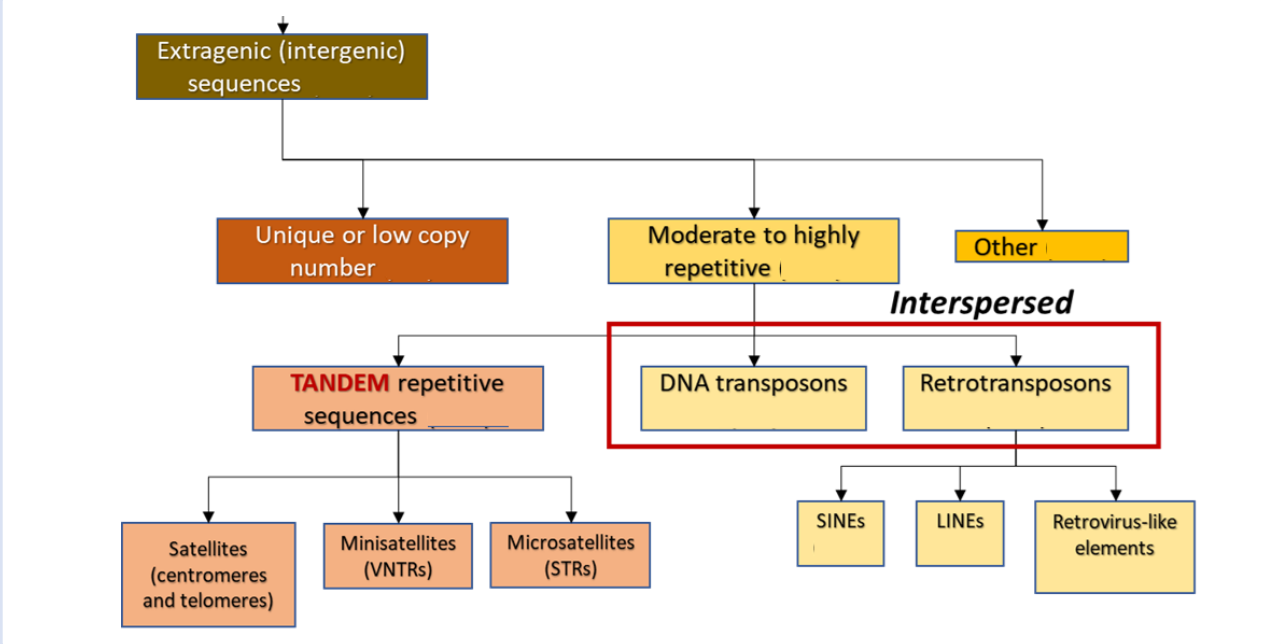


~5% of the genome contains sequences of noncoding DNA that are highly conserved (critical to survival).

Note: All numbers are approximate



- **Interspersed** means that they are dispersed repeats found in different regions of the human genome, distributed in the human genome all over the place in different chromosomes but are the basically the same sequences
- **Tandem** repeats means that they come one after the other, so it's the same sequence but it's repeated one after the other. Tandem means that they are linked & associated with each other



These are called transposons—motile sequences that can move around. Some originate from DNA viruses, some from RNA viruses. Around 40% of our genom originates from RNA viruses called retroviruses, and about 3% from DNA viruses. Retroviruses, like HIV, have RNA genomes that are reverse-transcribed into DNA before entering human DNA.

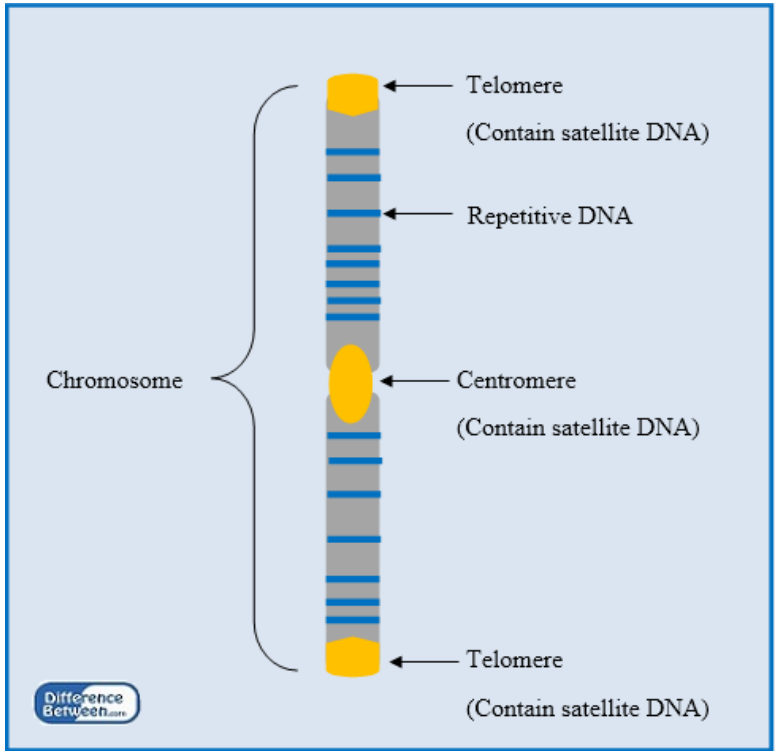
Tandem repeats

Satellite (macro-satellite) DNA

- Regions of 5-300 bp repeated 10^6-10^7 times
- **They are found in** Centromeres and telomeres
- **Every chromosome has it's own centromeric sequences but the telomeres have the same sequence**
- Centromeric A/T-rich repeats (171 bp) called α-satellite unique to each chromosome (you can make chromosome-specific probes) by fluorescence in situ hybridization (FISH).

For more explanation
Click here

Telo means end; telomeres exist at the ends of chromosomes

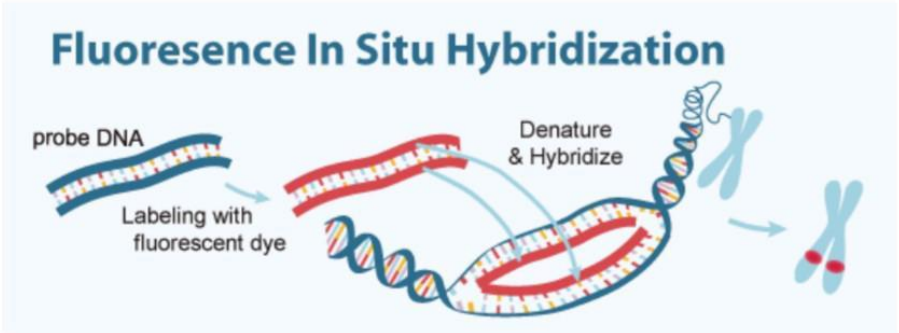


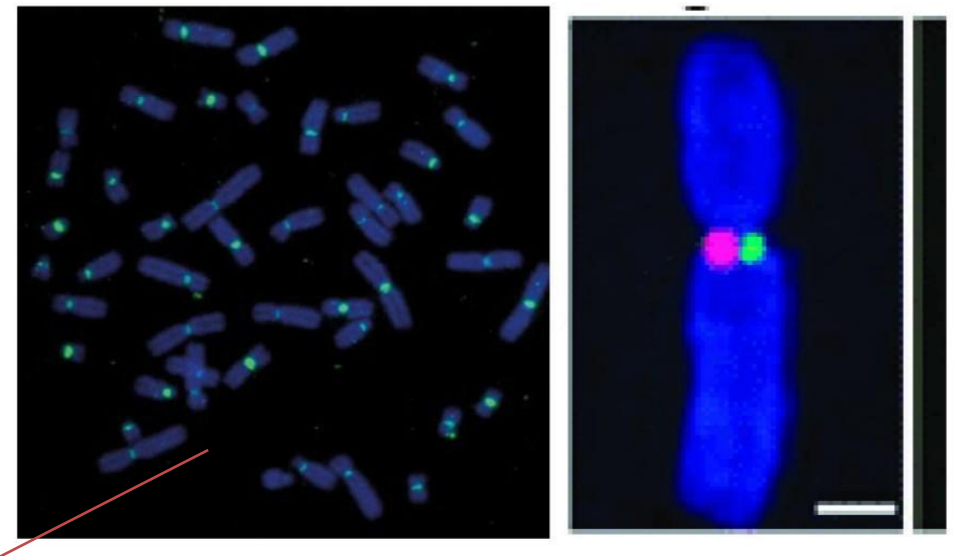
- Centromeres are regions that exist in the middle of chromosomes - not necessarily in the center- but they are the constricted regions in the middle of chromosomes and they divide chromosomes into Q arm and P arm.
- P stands for petite - small & short- and the Q arm is the long arm.

Fluorescence in situ hybridization (FISH)

In place

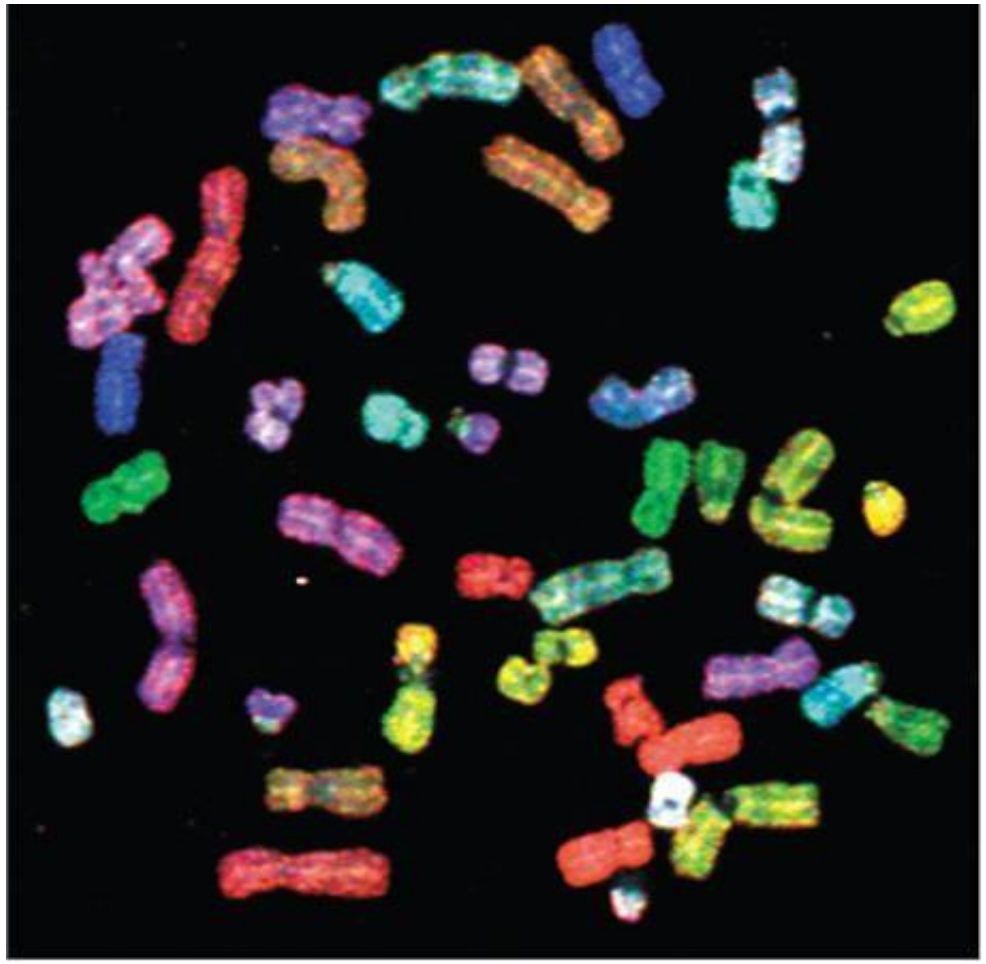
- FISH is a technique that uses fluorescently labeled oligonucleotide probes that are complementary to specific repeated sequences on individual chromosomes to visualize the location of chromosomes.
- Uses:
  - Locate a gene on a chromosome
  - Determine chromosomal and genetic anomalies like:
    - Duplication, deletion, translocation, and amplification





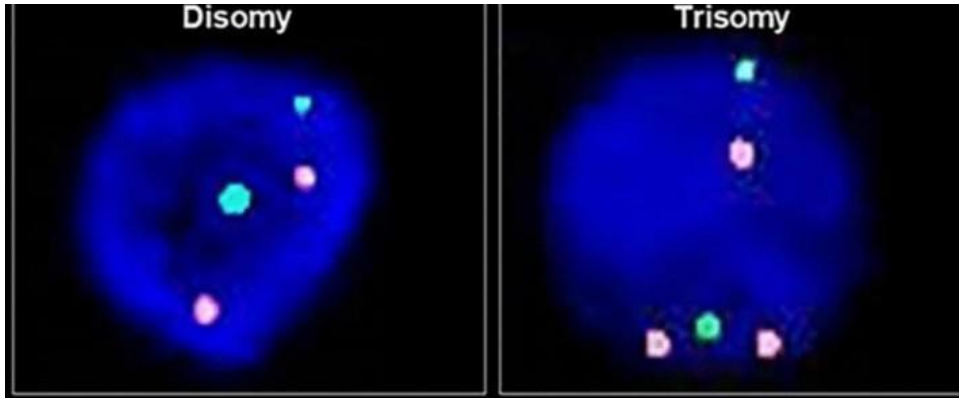
Using a technique known as fluorescence in situ hybridization, we can label different regions of the chromosome by using probes and the probes are labeled with a fluorescent tag or radioactivity -mostly fluorescent tags- and the probes would bind to chromosomes at certain regions.
You can label centromeres -as shown here- and you can label every single centromere since each chromosome has its own unique centromeric Repeat,,,, we can use 24 different probes, each one is labeled with a different color so we can label each chromosome with a different color.

Fluorescence in situ hybridization (FISH)

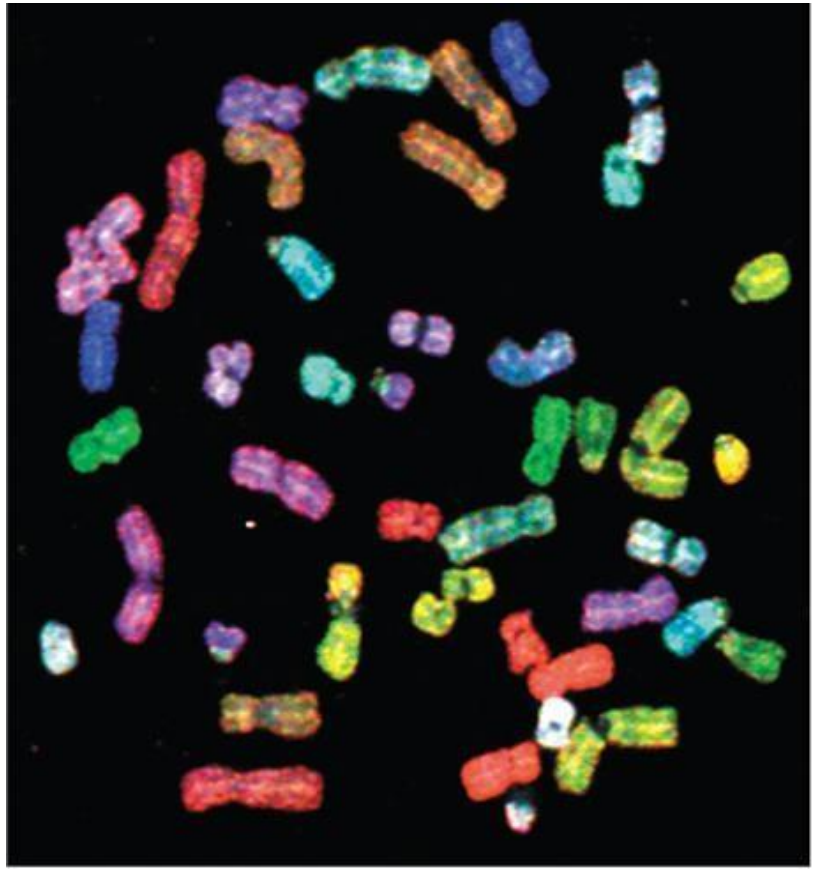


Courtesy of Thomas Ried and Hesed Padilla-Nash, National Cancer Institute

- Hybridization of human chromosomes with chromosome-specific fluorescent probes that label each of the chromosomes a different color.



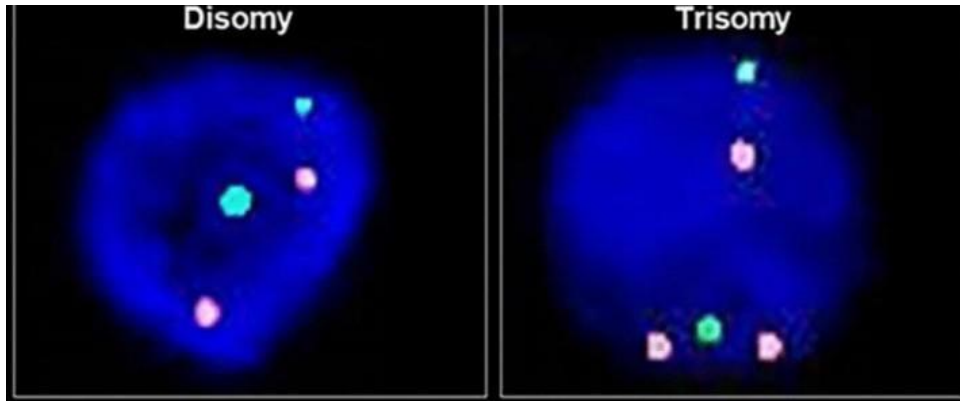
(I)



Courtesy of Thomas Ried and Hesed Padilla-Nash, National Cancer Institute

<u>We can use this technique to detect mutations at the chromosomal level.</u>
Notice in photo (I) that each chromosome is stained based on a sequence specific to that chromosome—for example, the centromere

(II) (III)



In photo (II), we stain two chromosomes using two different probes. Notice that because the cell is diploid and has two copies of each chromosome, two chromosomes are stained green and two are stained pink.

But in photo (III), we see three chromosomes that are pink. This means there is an extra copy of that chromosome. So with this method, we can detect whether there is an abnormal number of chromosomes or not, or if a deletion has occurred on a chromosome, or if an extra segment has been added to the chromosome, or not. We can also detect translocation—for example, when a segment from one chromosome moves to another chromosome, like chromosome 10

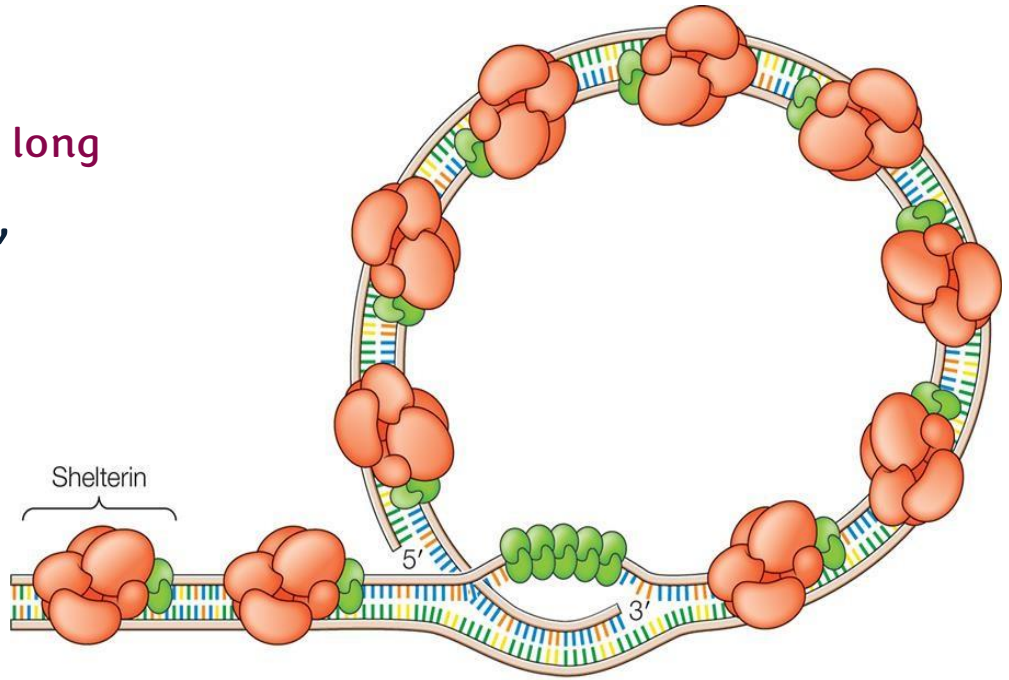
Telomeric repeats

They exist at the end of chromosomes

- (TTAGGG) is repeated hundreds to thousands of times at the termini of human chromosomes with a 3′ overhang of single-stranded DNA.
- The repeated sequences form loops that bind a protein complex called shelterin, which protects the chromosome termini from degradation and prevents fusion between chromosomes—just like tying a knot at the end of a thread to prevent it from unraveling.

- **Telomeric repeat-containing RNA (TERRA):** a long non-coding RNA transcribed from telomeres and functions in:
  - maintaining the integrity of chromosome termini,
  - regulating telomerase activity, This enzyme keeps telomeres long
  - maintaining the heterochromatic state of telomeres,
  - protecting DNA from deterioration or fusion with neighboring chromosomes

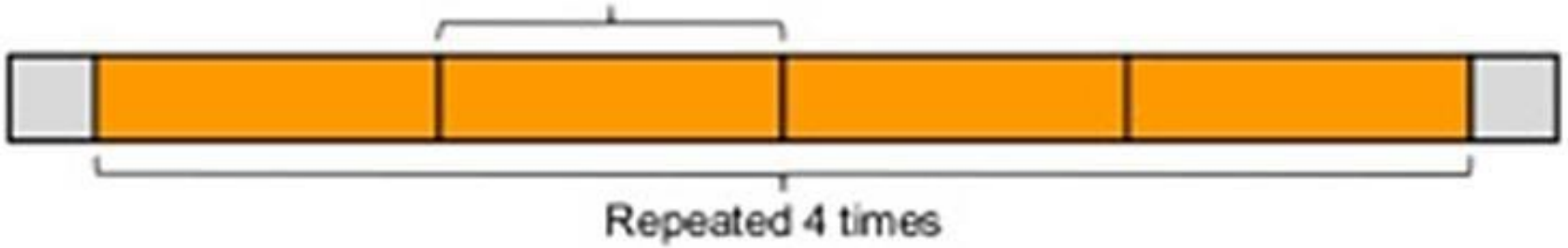
Both centromeres and telomeres are **heterochromatin**, which is tightly packed DNA. It was believed about them that they do not contain expressed genes. But we later found that they do produce RNA—specifically a long non-coding RNA called TERRA (Telomeric Repeat-Containing RNA)



Mini- and Micro-satellite DNA

Minisatellite: Variable Number Tandem Repeats (VNTR)

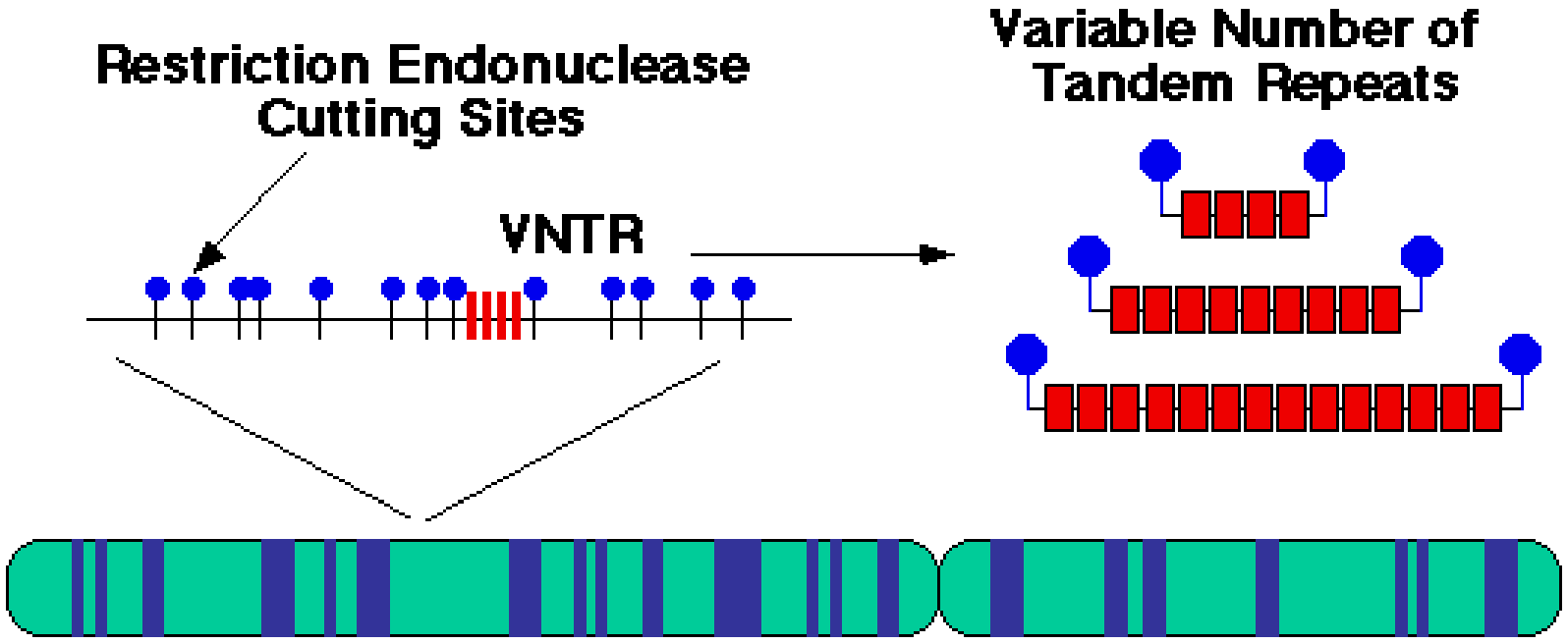
Tandem because one comes after the other; repeats because basically it's the same sequence repeated like several times.



Microsatellite: Short Tandem Repeats (STR) – Simple Sequence Repeats (SSR)

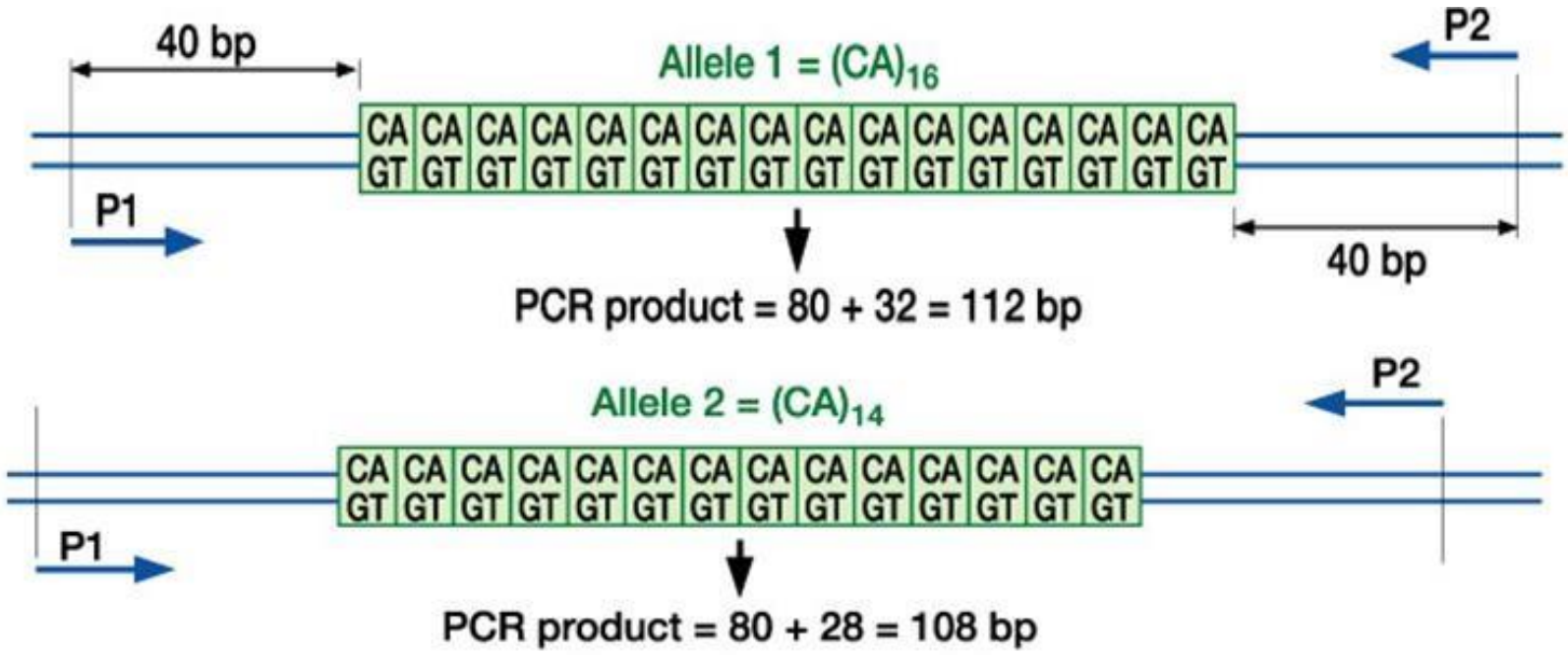
Mini-satellite DNA

- Mini satellite sequences or VNTRs (variable number of tandem repeats) of 20 to 100 bp repeated 20-50 times



Micro-satellite DNA

- STRs (short tandem repeats) of 2 to 10 bp repeated 10-100 times



Polymorphisms of VNTR and STR

By the way, if our DNA contains 3 billion nucleotides, then a 0.1% difference between individuals corresponds to about 3 million nucleotides – which is a substantial amount. These 3 million differences account for what we call genetic polymorphisms

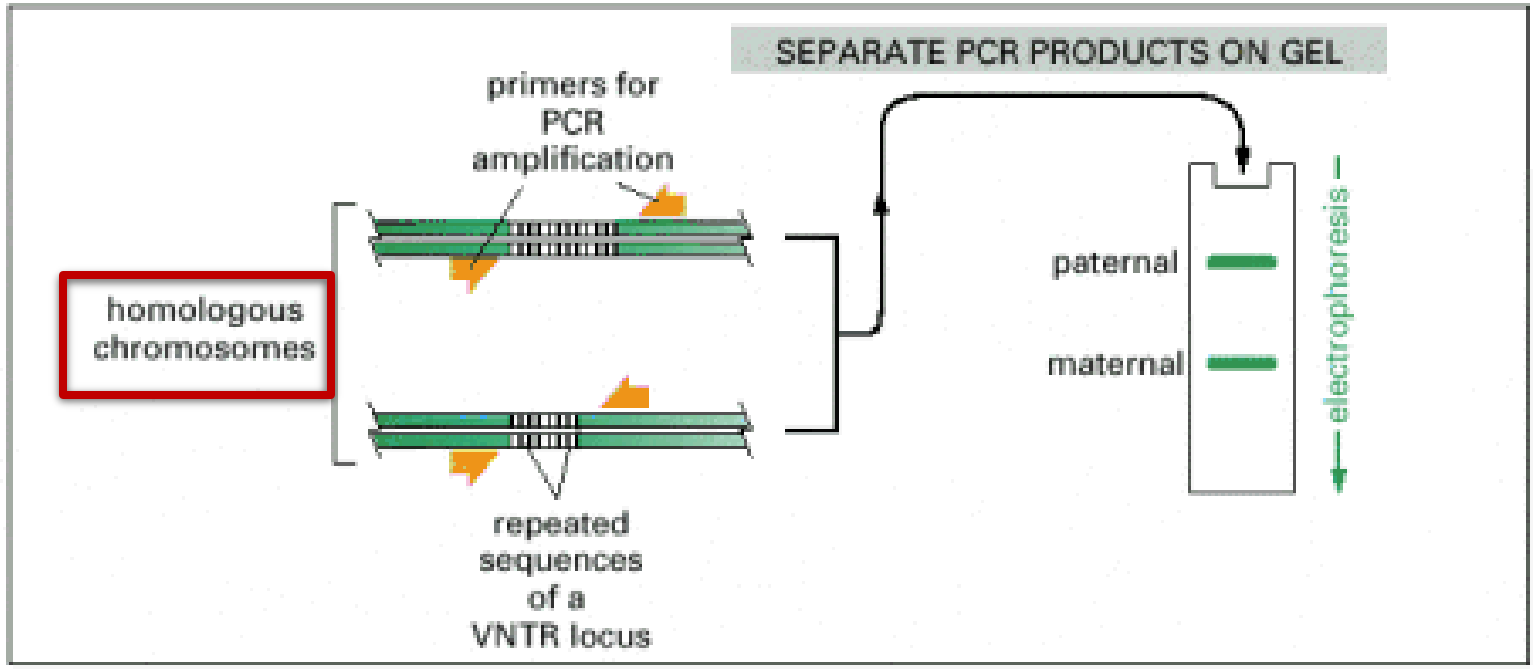
- STRs and VNTRs are highly variable among individuals (polymorphic).
  - They are useful in DNA profiling for forensic testing.

As individuals, we can have different numbers of repeated sequences. For example, on the paternal chromosome you might have an STR or VNTR repeated 10 or 20 times, while on the maternal chromosome the exact same repeat sequence—at the same chromosomal location—might be repeated 30 times.

So, essentially, by using RFLP (Restriction Fragment Length Polymorphism) or PCR, we can identify the polymorphisms present in our cells. These techniques allow us to determine how many repetitions of a VNTR or STR sequence exist in our DNA

Polymorphisms of VNTR and STR

- STRs and VNTRs are highly variable among individuals (polymorphic).
  - They are useful in DNA profiling for forensic testing.

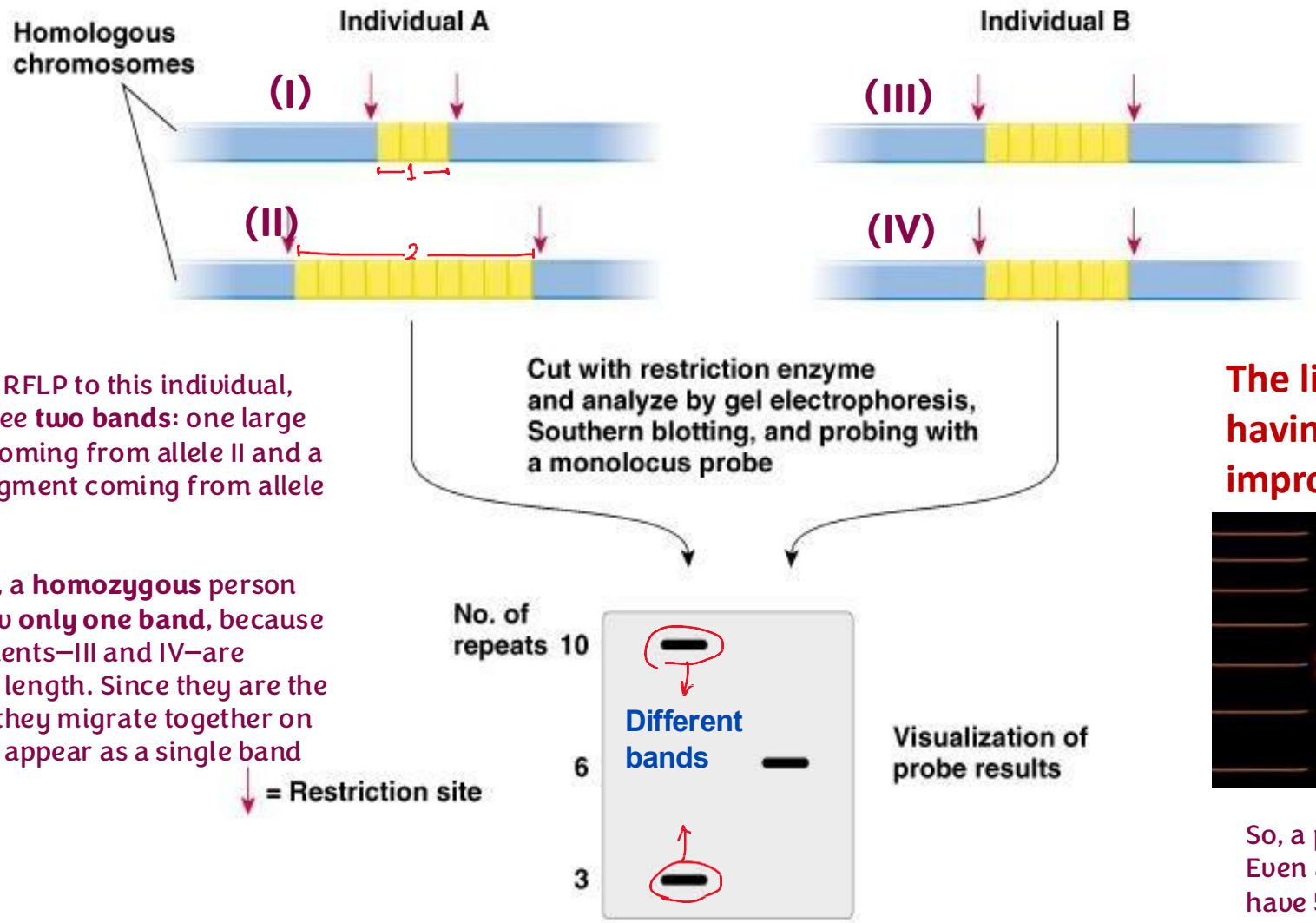


Homologous chromosome (or homologs) are the set of one maternal and one paternal chromosome somatic diploid cells.

Homologous chromosome: the same chromosome except that we get it one from mother and one from father .

STRs and VNTRs as DNA Markers

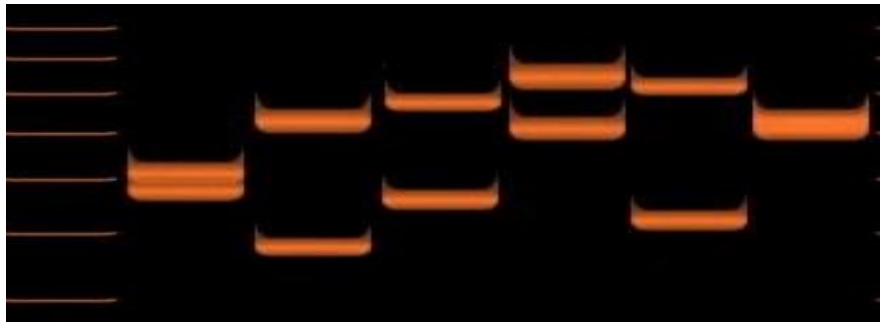
In **individual A** is heterozygous for a certain Str or vntr so basically if we cut the region 1and 2 we can have two different Bands . And in **individual B** a person can be homozygous for the vntr or the Str



If we apply RFLP to this individual, we would see **two bands**: one large fragment coming from allele II and a smaller fragment coming from allele I.

In contrast, a **homozygous** person would show **only one band**, because both fragments—III and IV—are identical in length. Since they are the same size, they migrate together on the gel and appear as a single band

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



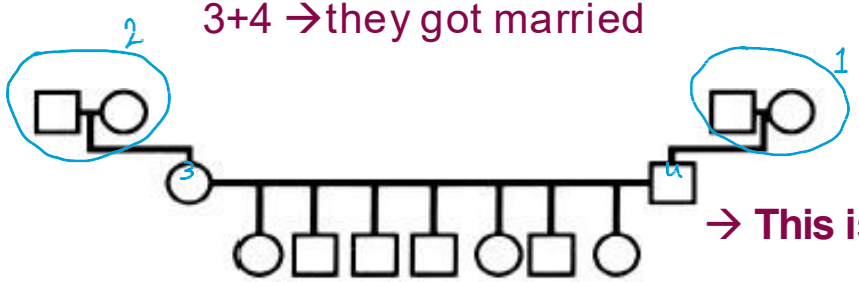
So, a person can be either heterozygous or homozygous. Even among homozygous individuals, one person might have 5 repeats on both chromosomes, while another has 10 repeats on both. Both are considered homozygous, but for different alleles. That's why we must be specific when we say someone is homozygous – homozygous for which allele or for which repeat number, or for which gene.

Note :When we examine a person's VNTRs and STRs, we compare these repeat patterns to the sample. If the VNTR and STR profiles match, then the sample is compatible with that individual. This is exactly what is done in criminal investigations, paternity testing, and the RFLP-based methods we discussed earlier

Real example

Note:The grandfather is represented by a square and the grandmother by a circle

1+2 → are grandparents
3+4 →they got married



→ This is a pedigree.

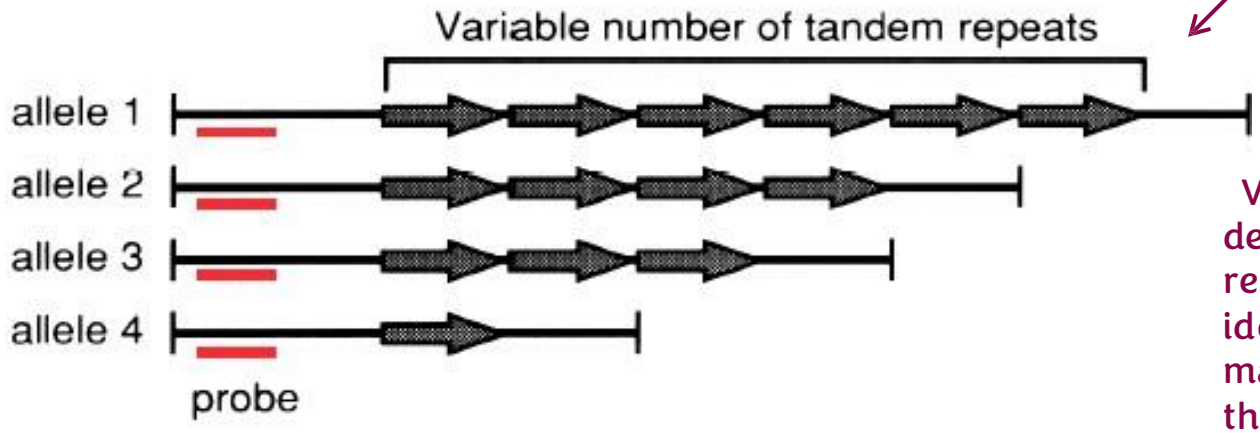
Large DNA fragment

Small DNA fragment

single-locus probe but multiple alleles

We cut the DNA and get and form different fragments like these over here

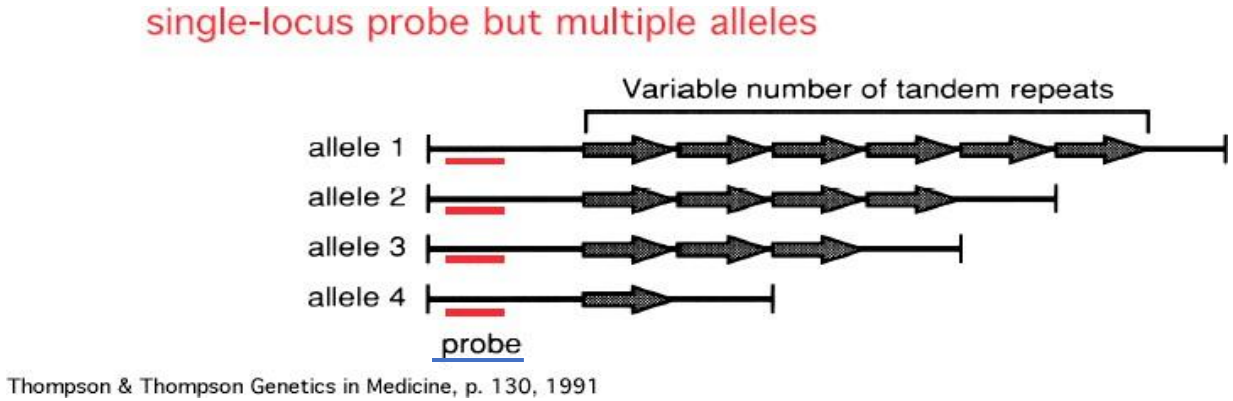
We separate DNA on a gel, then transfer the DNA form the gel to a membrane and add a probe.



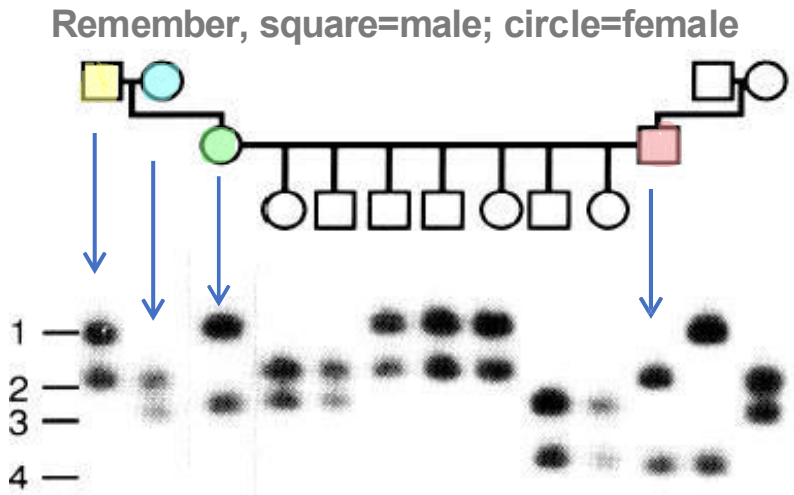
Variable length of fragments depending on number of repeats, other parts are identical in all fragments; main difference between them being number of repeats.

Thompson & Thompson Genetics in Medicine, p. 130, 1991

- Now, let's look at an example. Consider a pedigree of a family: male, female, grandparents, their daughter, and their son. The couple has seven children. Within this family, there is a specific VNTR present on for example chromosome 6. There are four alleles (Al=alleles)(Al1, Al2, Al3, Al4) with varying repeat numbers of VNTR: Al1 has six repeats, Al2 has four, Al3 has three, and Al4 has one repeat.
- If we use Southern blotting with a probe specific to this region ~the red line~ we can distinguish the alleles by their fragment lengths.
- We cannot use probe for VNTR because it would bind to all of them &we will not be able to distinguish different alleles, so we use probe that would bind outside VNTR region.



Thompson & Thompson Genetics in Medicine, p. 130, 1991



We are concerned with the size of fragments not necessarily intensity of the signal.

→For example, the grandfather has Al1 on the chromosome and Al2 on the other chromosome, while the grandmother has Al2 and Al3. Their daughter inherits Al1 from her father and Al3 from her mother. The son of the other family inherits Al2 from his mother and Al4 from his father. When these two marry and have children, each child's genotype reflects a combination of alleles from both parents (all of the kids here are heterozygous).

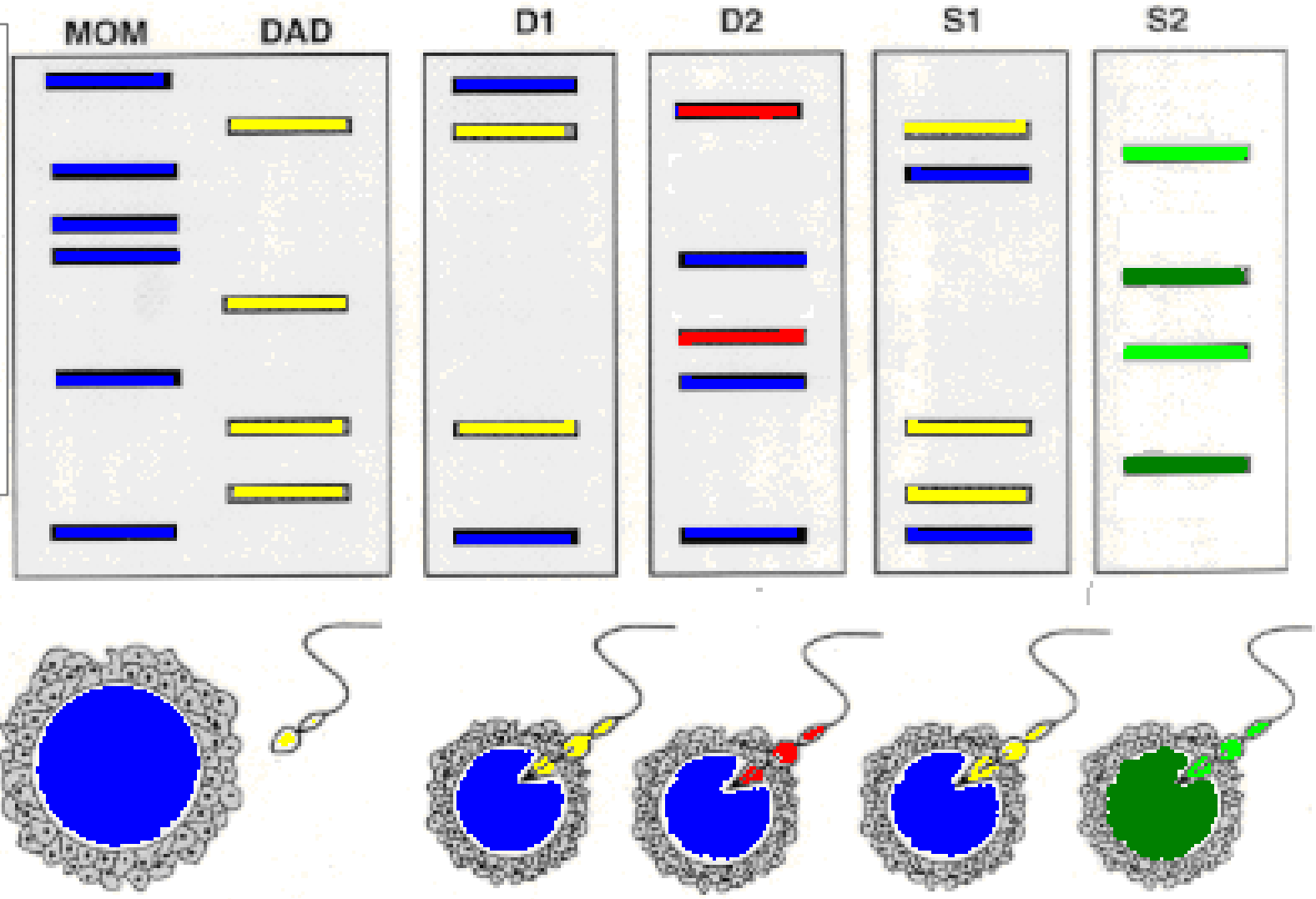
→We can see that each child has their own unique genotype, and while some may share a genotype -some genetic makeup being similar or identical in different individuals which is why we do several VNTRs to link a sample with a person- the majority are distinct.

Paternity testing

This is the genetic the molecular profile, also known as genetic profile, genetic fingerprinting & molecular fingerprinting.

- Each child has a fragment inherited from the mother and another from the father.
- D1& S1have fragments that can be linked to both parents genetic profile (actual children of both parents).
- D2 has fragments from the mother &no fragments from the dad.
- S2 is not the son of either parents.

If we observe a child whose genotype does not match either parent, it could indicate non- paternity (used in paternity testing).



Single nucleotide polymorphism (SNPs)

→All of the previously mentioned tandem repeats are variables, different among individuals.
→ Another source of genetic variation is single nucleotide polymorphisms (SNPs). SNPs are variations in a single nucleotide in the genome that occur at specific positions.

→Polymorphism means different shapes for single nucleotides

- Another source of genetic variation
- Single-nucleotide substitutions of one base for another
- Two or more versions of a sequence must each be present in at least one percent of the general population
- SNPs occur throughout the human genome - about one in every 300 nucleotide base pairs.
  - ~10 million SNPs within the 3-billion-nucleotide human genome
  - **Not all of them can give us information,,,just** Only 500,000 SNPs are thought to be relevant **(informative)**

Examples

Individual 1 heterozygous

Chr 2 ...CGATATTCC**T**ATCGAATGTC...
copy1 ...GCTATAAGG**A**TAGCTTACAG...

Chr 2 ...CGATATTCC**C**ATCGAATGTC...
copy2 ...GCTATAAGG**G**TAGCTTACAG...

Individual 2 homozygous

Chr 2 ...CGATATTCC**C**ATCGAATGTC...
copy1 ...GCTATAAGG**G**TAGCTTACAG...

Chr 2 ...CGATATTCC**C**ATCGAATGTC...
copy2 ...GCTATAAGG**G**TAGCTTACAG...

Individual 3 homozygous

Chr 2 ...CGATATTCC**T**ATCGAATGTC...
copy1 ...GCTATAAGG**A**TAGCTTACAG...

Chr 2 ...CGATATTCC**T**ATCGAATGTC...
copy2 ...GCTATAAGG**A**TAGCTTACAG...

Individual 4 heterozygous

Chr 2 ...CGATATTCC**T**ATCGAATGTC
copy1 ...GCTATAAGG**A**TAGCTTACAG

Chr 2 ...CGATATTCC**C**ATCGAATGTC
copy2 ...GCTATAAGG**G**TAGCTTACAG

Individual 5 heterozygous

Chr 2 ...CGATATTCC**C**ATCGAATGTC...
copy1 ...GCTATAAGG**G**TAGCTTACAG...

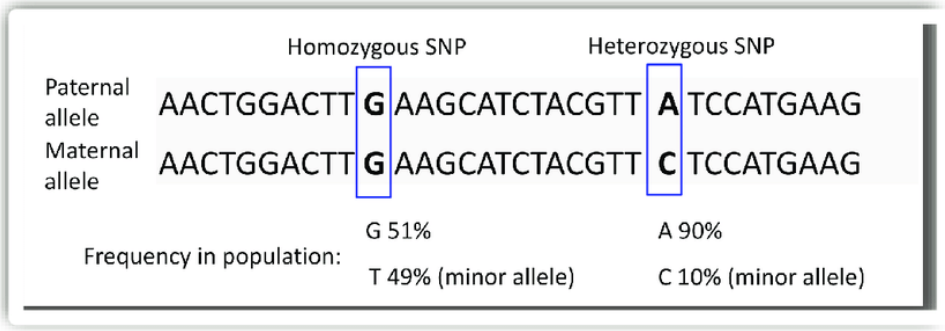
Chr 2 ...CGATATTCC**T**ATCGAATGTC...
copy2 ...GCTATAAGG**A**TAGCTTACAG...

Individual 6 heterozygous

Chr 2 ...CGATATTCC**C**ATCGAATGTC...
copy1 ...GCTATAAGG**G**TAGCTTACAG...

Chr 2 ...CGATATTCC**T**ATCGAATGTC...
copy2 ...GCTATAAGG**A**TAGCTTACAG...

This person is homozygous for a specific SNP (GG) & is heterozygous for another (AC).

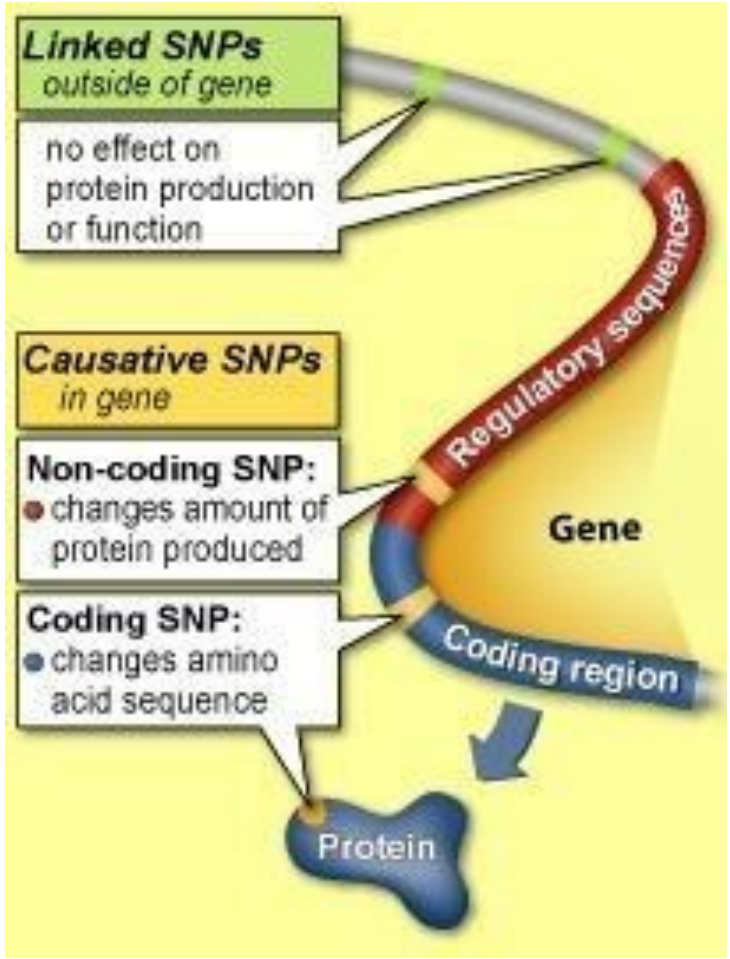


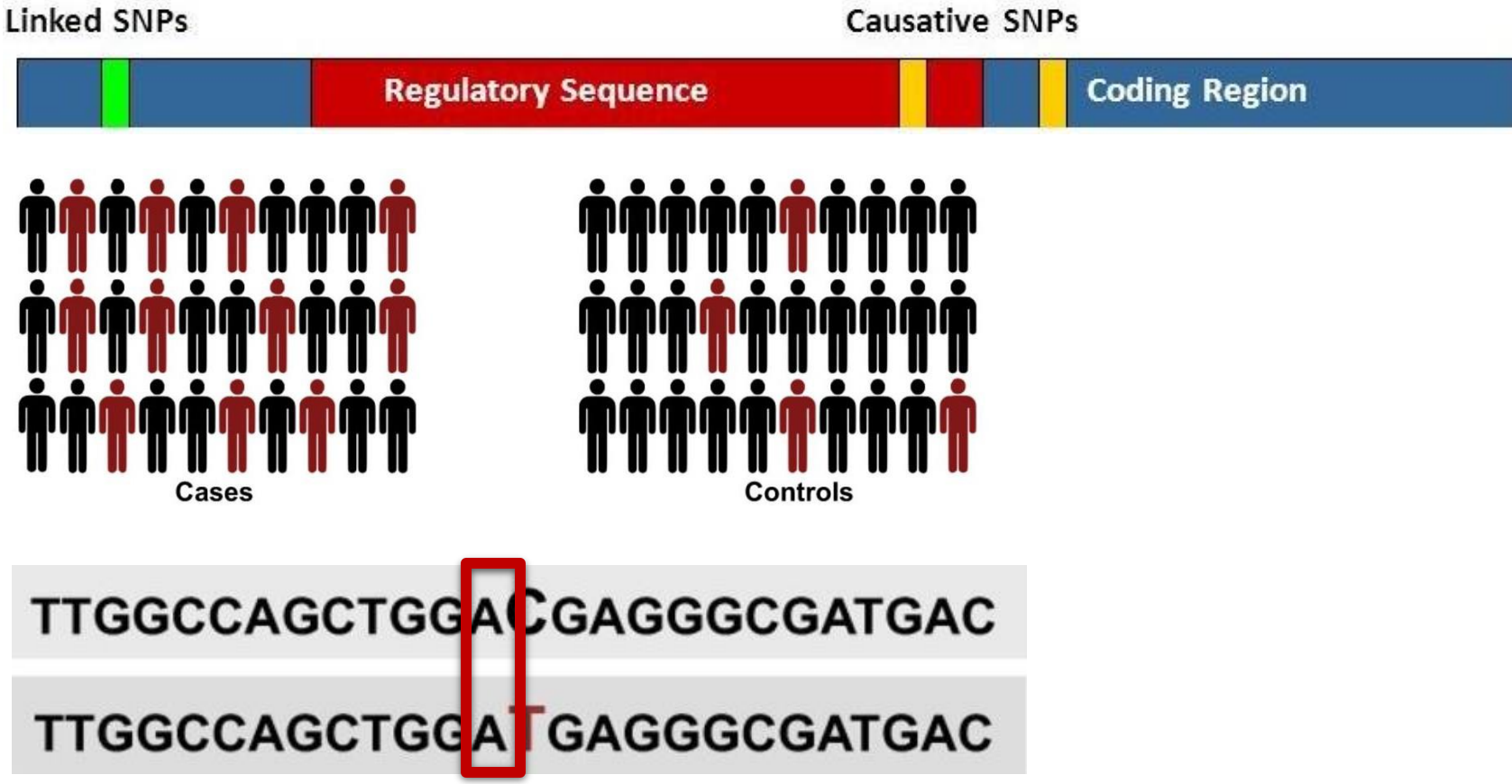
To be classified as a SNP, the variation must exist in more than 1% of the population; otherwise, it is considered a mutation (basically, a SNP is a type of mutation the exists in large proportion of a population).

SNPs can be homozygous or heterozygous, depending on whether the nucleotides at a specific position are identical
or different on homologous chromosomes.

Categories of SNPs

We have two types of SNPs that are relevant:
1)Linked SNPs (linked to a phenotype)
2) Causative SNPs (cause a phenotype)





- The significance of causative SNPs depends on their location.
- **SNPs in coding regions** can change the amino acid sequence of a protein(e.g. change from valine to alanine,,,both are non polar amino acid), potentially affecting its function(better or worse,, faster or slower).
- While those Non-coding SNPs in **regulatory regions** can regulate gene expression of the coding region - doesn't code for protein but regulates- and may cause less protein coded from coding region for example, which would affect person's health.
- However, most SNPs are linked SNPs that do not cause a phenotype -no functional significance- but are associated with phenotype . For example, person who has a diseases has a G in this part of the genome, doesn't mean causation it just indicates linkage for some particular reason. G basically acts as a marker for higher probability of disease affecting an individual.
- **What do we mean by causative SNPs?**

1. In one population, there would be variation between individuals based on different SNPs. For example, one person may require one pill of Panadol while another may require two....Why is this phenotype shown? Because there are SNPs, the enzymes that make the metabolism for paracetamol are different (People who respond to a single paracetamol tablet tend to metabolize the drug slowly, while those who need two tablets metabolize it more quickly)
2. When we say that there is a possibility that someone will have a certain disease, why? Because he has SNP, this SNP makes him susceptible, susceptible means that he is prepared, to having a disease, but it doesn't mean that he will have a disease for sure, no, he has a possibility of 30-40% that he will have a disease.

Interspersed repeats

Transposons (jumping genes)

Majority of our transposons have lost the ability to change places in our genome – but we have transposons in our DNA might able to move by chance (rare but it might happened)

- They are segments of DNA that can move from their original position in the genome to a new location.
- Two classes:
  - DNA transposons (3% of human genome)
  - RNA transposons or retrotransposons (42% of human genome).
    - Long interspersed elements (LINEs, 21%)
    - Short interspersed elements (SINEs, 13%)
      - An example is Alu (300 bp) Alu exists all over human genome & can be integrated into protein coding genes.
    - Retrovirus-like elements (8%) Genome is RNA not DNA, like HIV

How doe retrotransposons move and integrate?

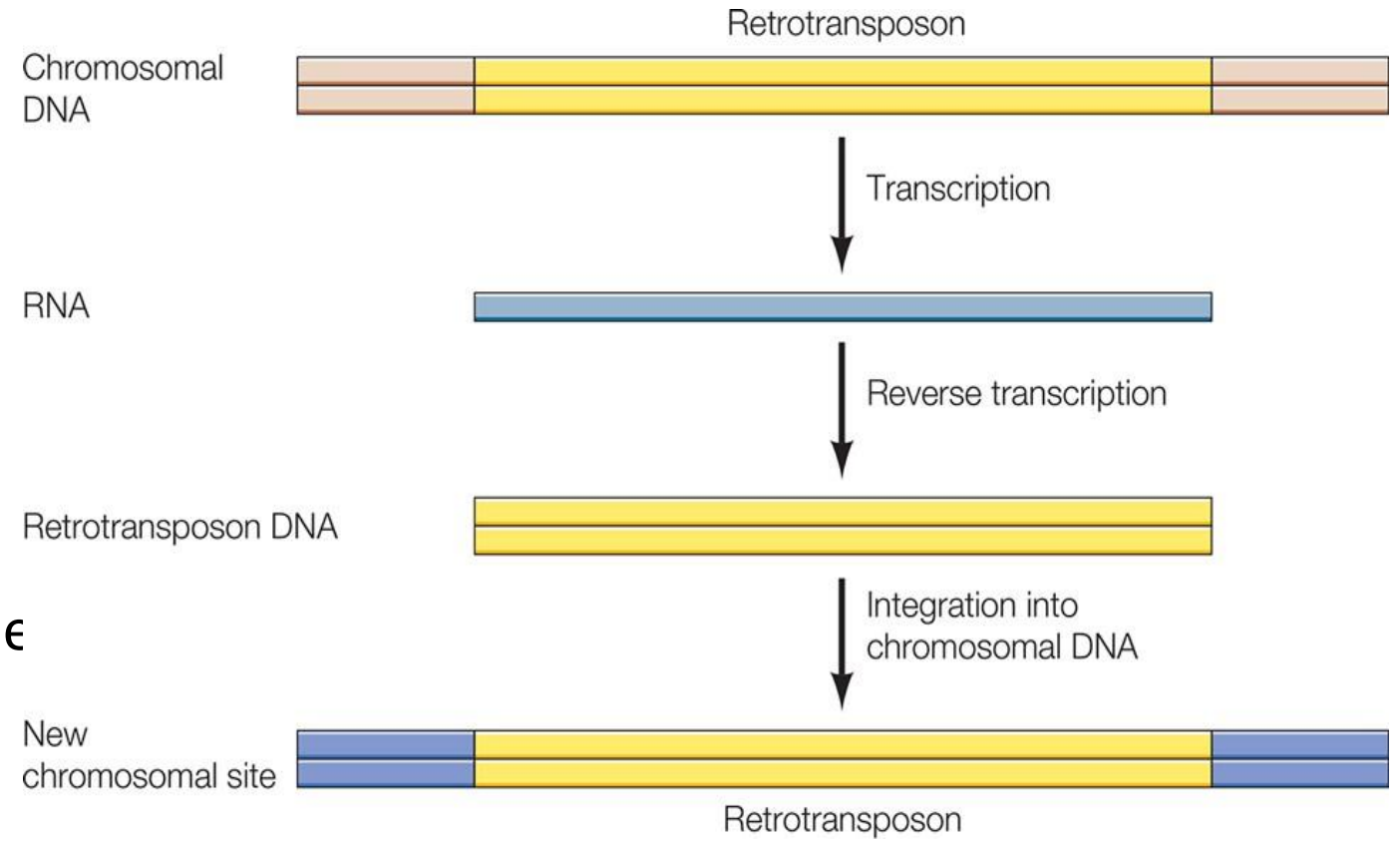
- A retrotransposon present at one site in chromosomal DNA is transcribed into RNA.
- The RNA is converted back into DNA by reverse transcriptase.

→Reverse transcriptase is an enzyme coded by some retrotransposons

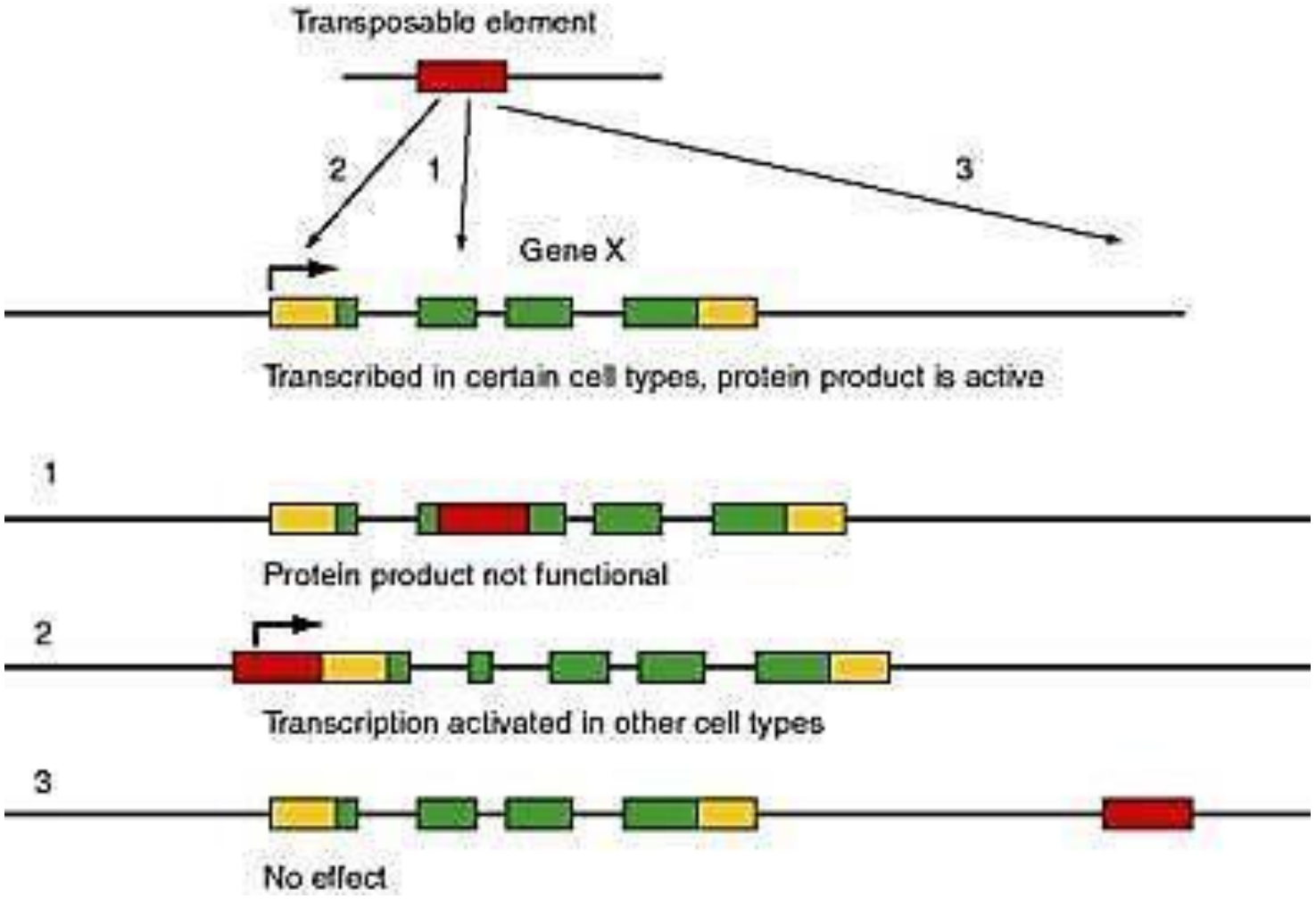
- The retrotransposon DNA can then integrate into a new chromosomal site.
- LINEs **(Long Interspersed Nuclear Elements)** contain <u>reverse transcriptase</u> genes and the <u>integrase</u> gene that is necessary for integration into cellular DNA.

→Integrase can also be coded by certain retrotransposons

→Majority of retrotransposons do not have reverse transcriptase & integrase gene, only some transposons can be transcribed & translated producing theses enzymes that help in there movement.



The outcome of transposition



- Over 99% of the transposons in the human genome lost their ability to move, but we still have some active transposable elements that can sometimes cause disease. (But in pigs they move, this is one of the problems in transplant)
  - Hemophilia A and B, severe combined immunodeficiency, porphyria, predisposition to cancer, and Duchenne muscular dystrophy. (These are diseases that can be caused by transposons)

The significance of transposons changing their place depends on where they move & get integrated

They can integrate:

1. Within gene making protein defective
2. Or integrate within regulatory sequence, like promoter or enhancer region, and affect activity of that gene
3. Outside of genes & regulatory sequences with no effect

Additional Resources:

رسالة من الفريق العلمي:

Reference Used:
(numbered in order as cited in the text)

1. DST modified
2. Chat gpt

يا خير مبعوثٍ وأفضلَ مُرسلٍ
يا من بمدحكَ تعجزُ الكلمات
صلى عليك اللهُ في ملكوتهِ
وعليك من رب الهدى البركاتُ

For any feedback, scan the code or click on it.

Corrections from previous versions:

Versions	Slide # and Place of Error	Before Correction	After Correction
V0 → V1	26	If we apply RFLP to this individual, we would see **two bands**: one large fragment coming from allele I and a smaller fragment coming from allele II.	If we apply RFLP to this individual, we would see **two bands**: one large fragment coming from allele II and a smaller fragment coming from allele I.
V1 → V2	16	106-107	10^6-10^7