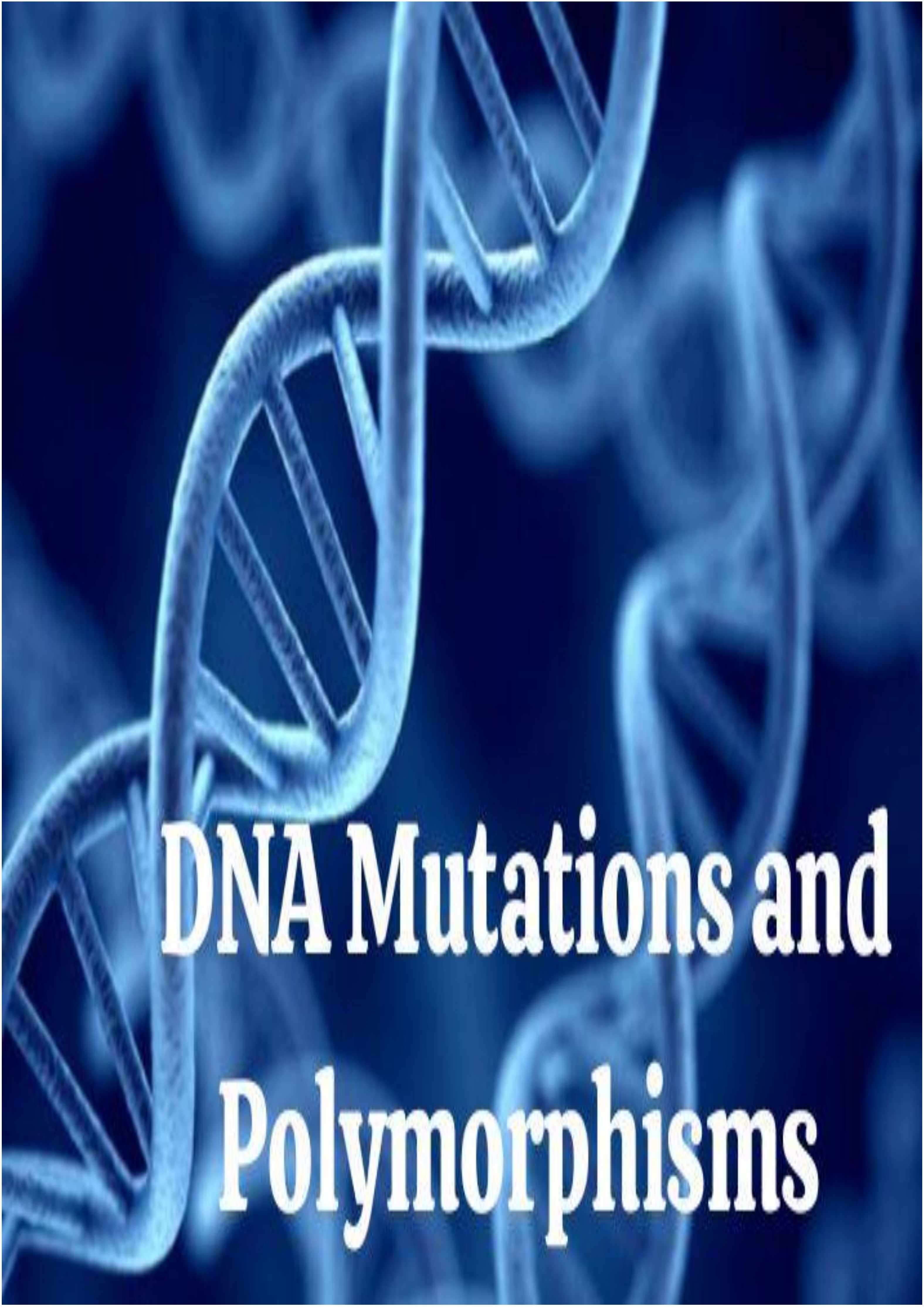


DNA Mutations and Polymorphisms

Mutations and polymorphisms

Locus : plural (Loci) , is a position of gene or base pair on chromosome.

Allele : an alternative form of gene.

Wild type : the more common allele in certain population

Variant : sometimes called **mutant,** the less common allele in certain population.

It might more than one variant exist in the same population

Mutation : permanent change in DNA sequence.

Polymorphism : common change in DNA sequence.

What is the main difference between **mutations** and **polymorphisms**?

Both involve change in DNA sequence, but the difference in the frequency of the change in the population, if the frequency of change < 1% we call it **mutation** ,but if the frequency of change > 1% we call it **Polymorphism.**

Some genes have more than one common Alleles(Multiple -forms) , and all forms are wild type , the highest frequency of them called **Common Variant.**

Mutations classified based on size into

- **Chromosome mutations** which involves added or remove whole chromosome.
- **Region or Sub regional chromosome mutations** , which involved add or remove part of the chromosome .
- **DNA mutation ,** which involve a change at DNA level.

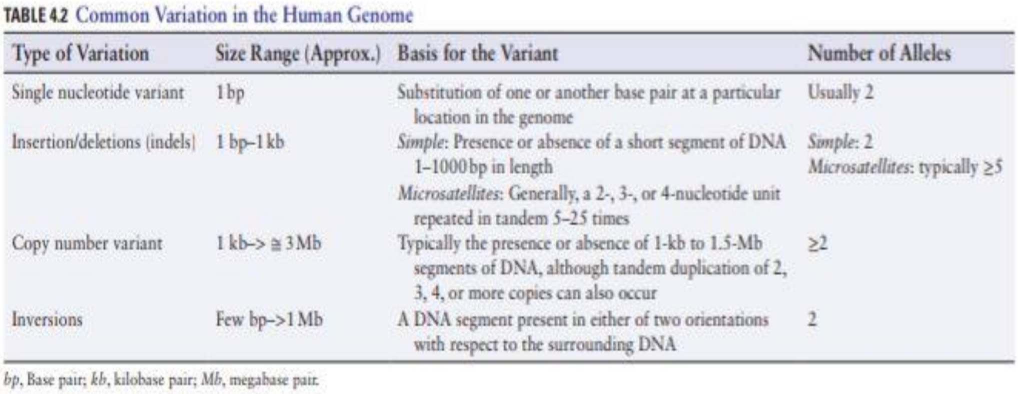


TABLE 4.2 Common Variation in the Human Genome

Type of Variation	Size Range (Approx.)	Basis for the Variant	Number of Alleles
Single nucleotide variant	1 bp	Substitution of one or another base pair at a particular location in the genome	Usually 2
Insertion/deletions (indels)	1 bp–1 kb	*Simple*: Presence or absence of a short segment of DNA 1–1000 bp in length *Microsatellites*: Generally, a 2-, 3-, or 4-nucleotide unit repeated in tandem 5–25 times	*Simple*: 2 *Microsatellites*: typically ≥5
Copy number variant	1 kb–> ≅ 3 Mb	Typically the presence or absence of 1-kb to 1.5-Mb segments of DNA, although tandem duplication of 2, 3, 4, or more copies can also occur	≥2
Inversions	Few bp–>1 Mb	A DNA segment present in either of two orientations with respect to the surrounding DNA	2

bp, Base pair; *kb*, kilobase pair; *Mb*, megabase pair.

Types of variations in human genome

- **SNPs (single nucleotide polymorphism)**

Which involved change only in one nucleotide, which is similar to **point mutation** which also involve a change of one nucleotide to another nucleotide but the main difference between the **SNP** and the **point of mutation** is the frequency of change it is more than one percent or less than one percent.

Like a point mutation, **SNPs could be have no effect on protein structure, it could be affect the protein activity and give non-functional protein or it could be lead to abnormal protein structure, or it could be have advantages for proteins.**

- **Indels (Insertion- Deletion mutations)**

  Which involve addition or deletion of base Pairs from DNA Sequence , it could be Deletion or insertion a few base pairs, or in certain cases A huge fragment Deleted or inserted.

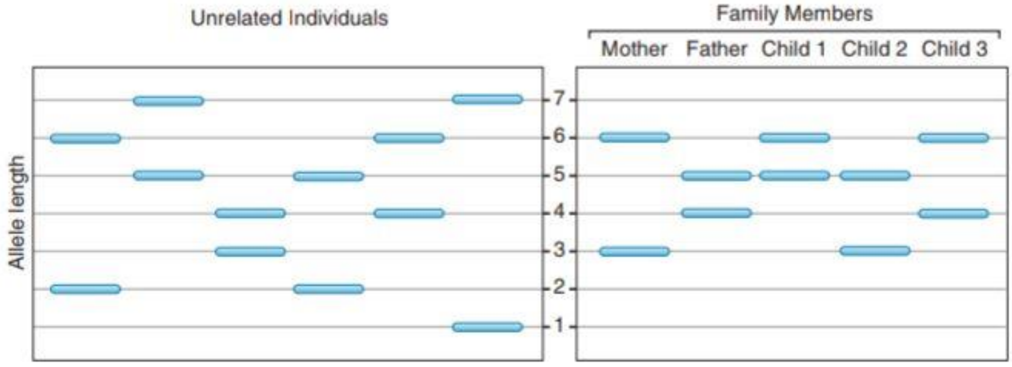
  Micro satellites: it is considered of **short tandem repeats** which generally involve di- tri -Tetra a nucleotide that were repeated many times.

  There are two types of repeats in human genome

  1. **Short Tandem Repeats** : Generally less than 10 nucleotides were repeated many times , for example **GCC GCC GCC GCC GCC GCC .**

  2. **Variable number repeats:** generally more than 10 nucleotides were repeated many times , in some cases **thousands** of base Paris were repeated many times .

     Micro satellites have been used in forensic science and paternity analysis .



This figure show using of micro satellites in paternity analysis, the child must be 50% compatible with father and 50% compatible .

Each band represent allele , remember child was inherited one Allele from his father and the second allele from his mother .

- **Inversion**

A segment of DNA was cut and flop and rejoined which lead to **inverted DNA fragment,** this could be happened in biological processes called **site specific recombination ,** which involves a set of proteins and factors include **Recombinase enzyme .**

Some Bacteria species use this mechanism to regulate Gene expressions by express two types of **flagellin** protein to escape from **immune system of host Organism .**

- **Copy number**

  how many number of this Chromosome /gene / DNA segment in the Cell ?

  The answer depend on what is the Gene or DNA you are looking for , for example in our Genome most of genes exist in **two copy, which mean the copy number of this gene is Two , there are a group of genes present on X Chromosome only , the Copy number of these genes is different from males to females , also there are some disorders like down syndrome there are three copies chromosome 21 , which mean there is three copies of these genes (copy Number = 3) ,if we look to DNA fragment it might present in four copies In individual 2 and in 6 copies in individual 2.**

- **Mobile elements (transposable elements)**

Repeated sequences that can “move” from one place in the genome to another by process called **transposition**

the Element moves to a new position in the genome either :

1. **Cut and paste mechanism**

Element migrate to new site without leave a copy in the original site

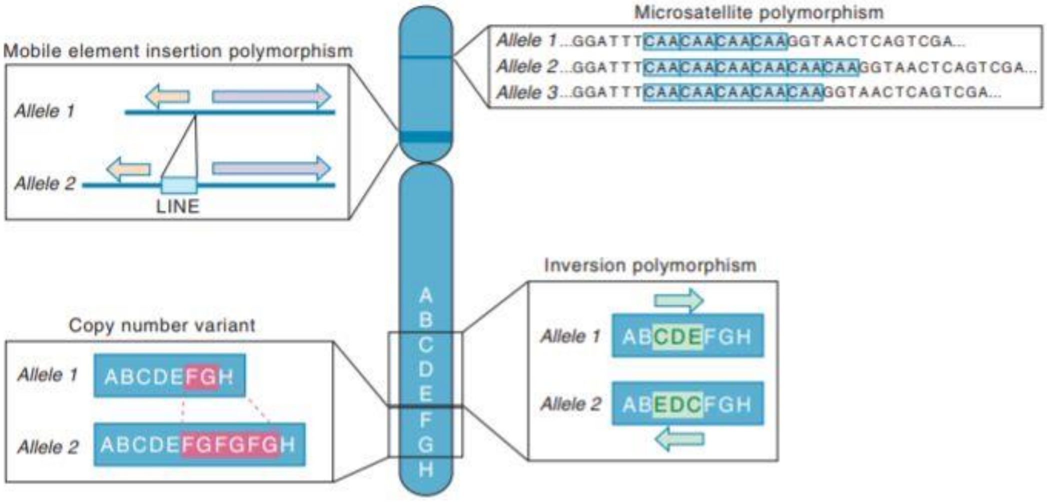
2. **Copy and paste**

Transposon leave a Copy in original site when it migrate to new site .

ALU element and LINE are example of Mobile elements

There is a type of transposition **called retro transposition**

Some viruses like HIV virus is **RNA virus ,** when the HIV enter the host cell to allow the Replication of HIV genome it must be convert from RNA to DNA and use host cell enzymes, special enzyme encoded by HIV virus called reverse transcriptase which convert RNA genome into DNA to allow Replication of virus genome, but within the cytoplasm there are many Types of RNA rather than Virus Genome , Reverse Transcriptase could by return the m-RNA molecules into DNA and Allow them to integrate into host genome .


Microsatellite polymorphism
Allele 1 ...GGATTTCAACAACAACAAGGTAACTCAGTCGA...
Allele 2 ...GGATTTCAACAACAACAACAACAAGGTAACTCAGTCGA...
Allele 3 ...GGATTTCAACAACAACAACAAGGTAACTCAGTCGA...
Mobile element insertion polymorphism
Allele 1
Allele 2
LINE
A
B
C
D
E
F
G
H
Inversion polymorphism
Allele 1
ABCDEFGH
Allele 2
ABEDCFGH
Copy number variant
Allele 1
ABCDEFGH
Allele 2
ABCDEFGFGFGH

Reference sequence

The Common DNA sequence in population, reference sequence is different from population to another population.

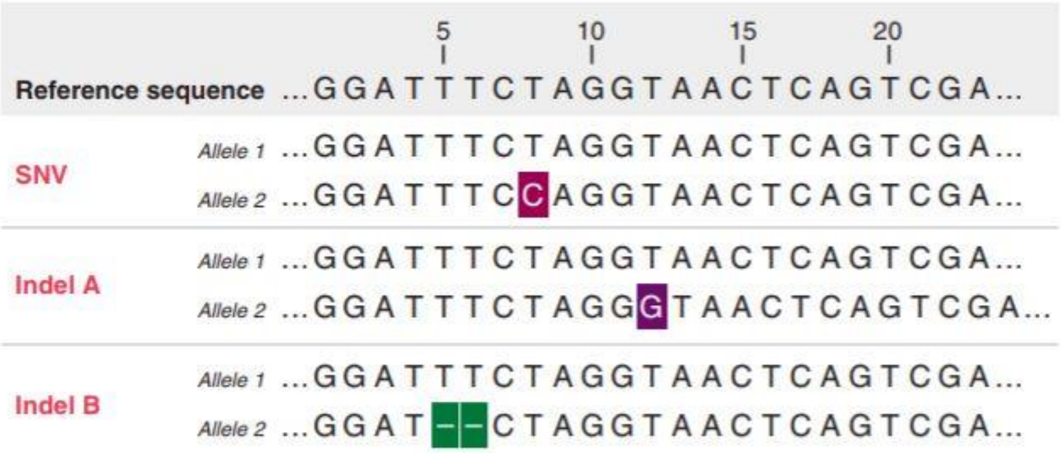
All DNA sequences that a present on databases are reference sequences which are used to compare with DNA sample for Patients to determine if has a Mutation/ SNP or not .

Reference sequence is **control** , the DNA sequence want to test it called **Query .**

It's not necessary all the people in population have identical sequences to reference sequence but they are similar to reference sequence for their population.

There are many techniques have used to detect different types of mutations and polymorphisms the most powerful technique is **whole genome sequencing** .

- Isolation of DNA sample from the patient
- Whole Genome Sequencing
- Finally compare Query with control



Reference sequence		...GGATTTCTAGGTAACTCAGTCGA...
SNV	Allele 1	...GGATTTCTAGGTAACTCAGTCGA...
	Allele 2	...GGATTTCCAGGTAACTCAGTCGA...
Indel A	Allele 1	...GGATTTCTAGGTAACTCAGTCGA...
	Allele 2	...GGATTTCTAGGGTAACTCAGTCGA...
Indel B	Allele 1	...GGATTTCTAGGTAACTCAGTCGA...
	Allele 2	...GGAT--CTAGGTAACTCAGTCGA...

Germline mutation : occurs in germline cells which responsible to produce gametes (sperm and egg) , this type can transmit into offspring, **all cells in child will contain the mutation.**

Somatic mutation : occurs in somatic cells and doesn't transmit into offspring.

Chromosome mutation: mis-segregation of chromosomes during Mitosis or Mitosis, also the effect of mutation depend on at where and when the mutation was happened , was it happened at embryonic stage or adult stage ? If it was happened at embryonic stage , was it happened at early stage or late stage ? .

If the mutation was happened at zygote stage the affect of mutation will be more severe .

One per 25-25 mitotic cell divisions , but this is under estimation because of Abortion before clinical diagnosis.

Regional mutations (During repair process)

Deletion or insertion for part of chromosome either during

- **Homologous recombination (Deletion, insertion, inversion)**
- **None homologous end joining repair (translocation, inversion).**

Gene mutations

- DNA Replication errors
- DNA repair errors

DNA Replication errors

DNA polymerase is the Enzyme which has crucial rule in DNA Replication by adding deoxy-nucleotides into growing chain to produce the new DNA strand , some time DNA polymerase add incorrect nucleotide into growing chain , this mis match if doesn't repair before the reach the second cycle of DNA replication it become permanent changed (Mutation) , DNA polymerase has proofreading activity which allow DNA polymerase repair the mis matching before reach second cycle of Replication.

DNA polymerase introduce 1 error every 10^7

Proofreading activity correct 99.9% of errors

From 10^ 3 incorrect nucleotide one incorrect nucleotide escape from proofreading.

By multiply the two numbers = 1× 10^10

The rate of mutation is 1 out 10^10 which equal

1÷ 10^10= 1×10^-10

DNA repair mutation

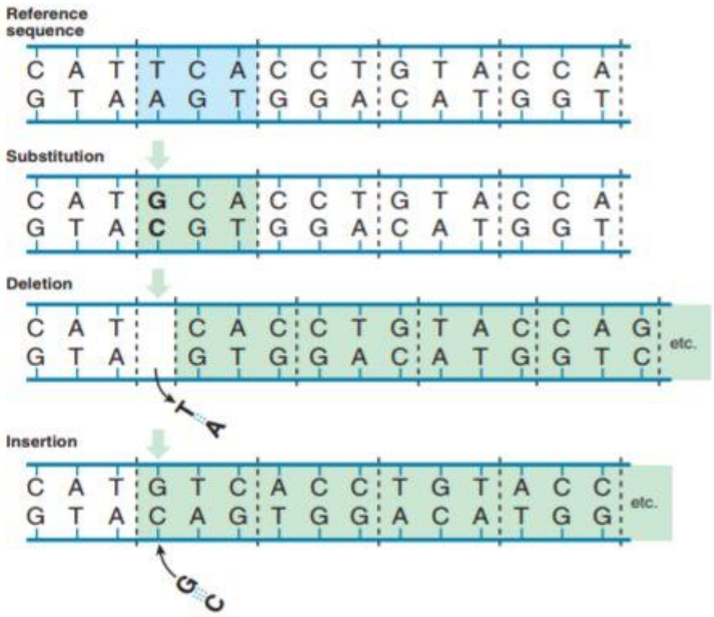
Mutagen : chemical, physical, or biological agent that could be

Induce change in DNA sequence.

Mutagen could be natural (**X-ray and UV) ,** of it could be manufactured (**detergents, cigarette) .**

Mutagens may case **deamination** (remove amino group) , **demethylation (remove methyl group) , and de-purination.**

The most common spontaneous mutation is convert c to T deamination of methyl cytosine result in thymine .


Reference sequence
C A T T C A C C T G T A C C A
G T A A G T G G A C A T G G T
Substitution
C A T G C A C C T G T A C C A
G T A C G T G G A C A T G G T
Deletion
C A T C A C C T G T A C C A G
G T A G T G G A C A T G G T C
etc.
T A
Insertion
C A T G T C A C C T G T A C C
G T A C A G T G G A C A T G G
etc.
G C

DNA mutation rate

Determine by whole genome Sequencing of trios (**Father , Mother and son)**

The rate of new mutations average between paternal and maternal gametes equal **1.2 × 10^8** mutations per base pair per generation.

Zygote require **2 gametes one paternal and the other maternal.**

There are 3 × 10^9 base pair but because the human is Diploid organism , this mean = 6 × 10^9 base pair

Each gamete contains 3×10^9 base pair

By multiply the number of base pairs in human with the rate of new mutation we can predict number of new mutations that presence in Child and never exist in his parents

6× 10^9 × 1.2×10^-8 = 72 (approximately)

We can estimate **rate of disease causing mutation per locus per generation** is used to measure new incidence of new cases of genetic disease not present in either parents.

Achondroplasia: Autosomal dominant disease causes by mutation in fibroblast Growth factor Receptor III , the patient is shorter than normal individuals, also there is no change in intellectual ability.

For example in certain hospital there is 244257 new born , 7 of them have Achondroplasia , after diagnosis of parents for affected babies they were found the parents were normal, how ?

The most explanation that affected babies were received the mutation during **gametogenesis**, the parents were normal but one of the parents during gametogenesis mutation was occurred in fibroblast growth factor Receptor III , and the Gamete that was carried the mutation entered in fertilization.

7 ÷ 242275 = 1.4× 10^-5 **rate disease causing mutation per locus per generation for Achondroplasia.**

The median gene mutation rate **1× 10^-6**

At least 5000 gene in human genome ,in which mutations cause disease.

5000 × 1×10^-6 = 5/1000

=**1/200 ,** is the probability of person to receive a new mutation in a known disease associated gene from one or the other parents .

Sex and age affect on mutation rate .

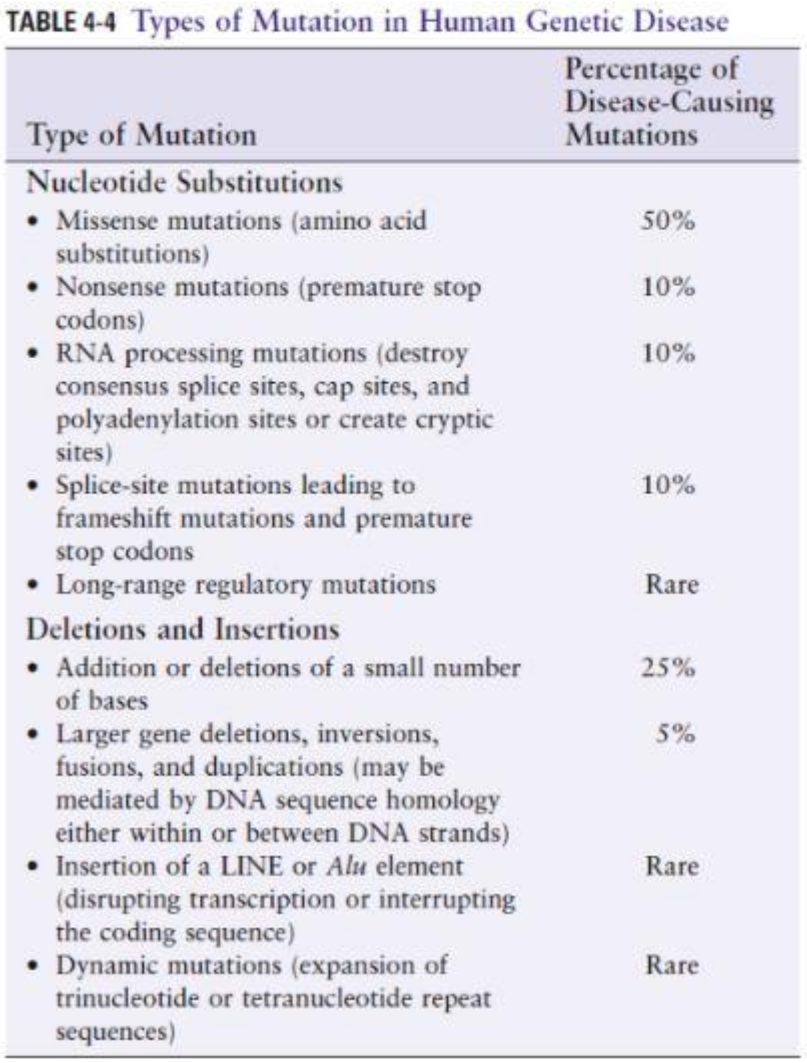


TABLE 4-4 Types of Mutation in Human Genetic Disease

Type of Mutation	Percentage of Disease-Causing Mutations
Nucleotide Substitutions	
• Missense mutations (amino acid substitutions)	50%
• Nonsense mutations (premature stop codons)	10%
• RNA processing mutations (destroy consensus splice sites, cap sites, and polyadenylation sites or create cryptic sites)	10%
• Splice-site mutations leading to frameshift mutations and premature stop codons	10%
• Long-range regulatory mutations	Rare
Deletions and Insertions	
• Addition or deletions of a small number of bases	25%
• Larger gene deletions, inversions, fusions, and duplications (may be mediated by DNA sequence homology either within or between DNA strands)	5%
• Insertion of a LINE or *Alu* element (disrupting transcription or interrupting the coding sequence)	Rare
• Dynamic mutations (expansion of trinucleotide or tetranucleotide repeat sequences)	Rare

The most common of mutations are **missense mutations**

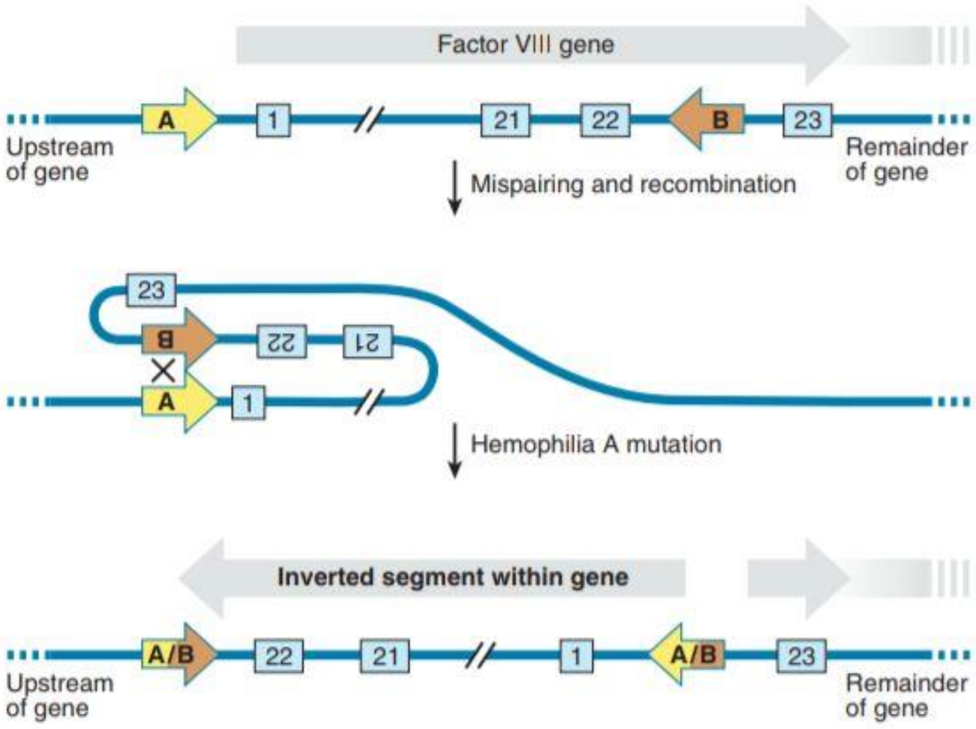
Missense mutation : change amino acid to another amino acid

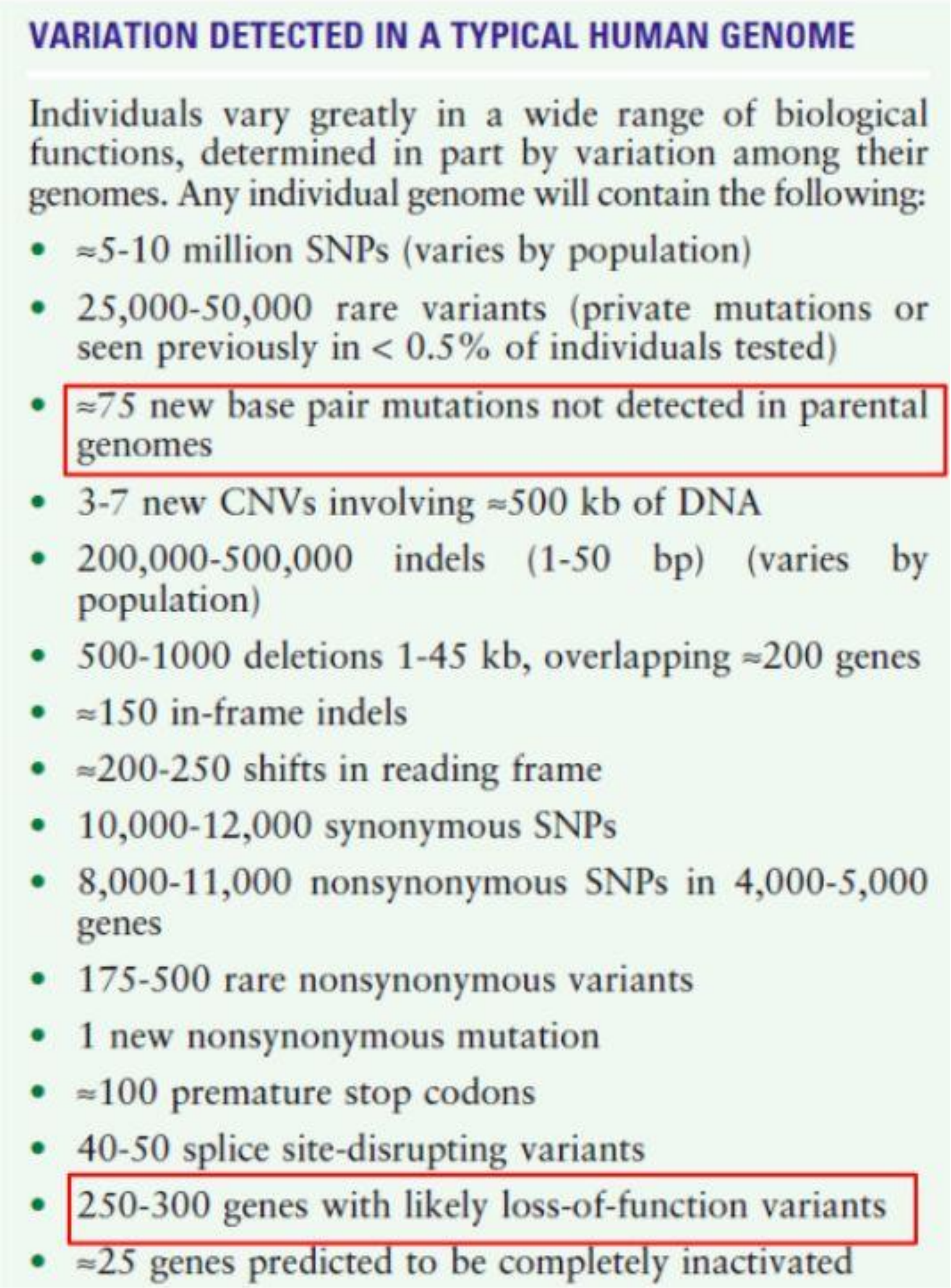
Nonsense mutation: premature stop codon (the most dangerous because usually the protein loss it function) the consequence depend on where the premature codon present , is it a the end or at beginning?

Other types of mutation affect the splicing sites which disrupt the splicing machinery

Inversion mutations

For factor VIII gene there are two homozygous Sequences **A and B apart 500 Kb from each other , but inverted ,** we Already know the chromosomes are interact with other and also the same chromosome fold in different patterns for certain purpose , on some cases **mis- intra-chromosomal interaction occur which lead to fold the X chromosome in the certain way in which the inverted Regions have the same direction at that moment , special enzyme recognize Region A and B and cut the DNA within both and re join it in inverted orientation.**





VARIATION DETECTED IN A TYPICAL HUMAN GENOME

Individuals vary greatly in a wide range of biological functions, determined in part by variation among their genomes. Any individual genome will contain the following:

- ≈5-10 million SNPs (varies by population)
- 25,000-50,000 rare variants (private mutations or seen previously in < 0.5% of individuals tested)
- ≈75 new base pair mutations not detected in parental genomes
- 3-7 new CNVs involving ≈500 kb of DNA
- 200,000-500,000 indels (1-50 bp) (varies by population)
- 500-1000 deletions 1-45 kb, overlapping ≈200 genes
- ≈150 in-frame indels
- ≈200-250 shifts in reading frame
- 10,000-12,000 synonymous SNPs
- 8,000-11,000 nonsynonymous SNPs in 4,000-5,000 genes
- 175-500 rare nonsynonymous variants
- 1 new nonsynonymous mutation
- ≈100 premature stop codons
- 40-50 splice site-disrupting variants
- 250-300 genes with likely loss-of-function variants
- ≈25 genes predicted to be completely inactivated

Done by :

Mohammed Qandeel