**Genetic marker**

A genetic marker is a gene or [DNA sequence](https://en.wikipedia.org/wiki/DNA_sequence) with a known location on a [chromosome](https://en.wikipedia.org/wiki/Chromosome) that can be used to identify individuals or [species](https://en.wikipedia.org/wiki/Species). It can be described as a variation (which may arise due to mutation or alteration in the genomic loci) that can be observed. A genetic marker may be a short DNA sequence, such as a sequence surrounding a single base-pair change ([single nucleotide polymorphism](https://en.wikipedia.org/wiki/Single_nucleotide_polymorphism), SNP), or a long one, like [minisatellites](https://en.wikipedia.org/wiki/Minisatellite)

**Single-Nucleotide Polymorphism**

SNPs are variations in a DNA sequence that occur when a single nucleotide in the sequence is different from the norm in at least one percent of the population. When SNPs occur inside a gene, they create different variants, or alleles, of that gene.



For example, at a specific base position in the human genome, the base C may appear in most individuals, but in a minority of individuals, the position is occupied by base A. There is a SNP at this specific base position, and the two possible nucleotide variations - C or A - are said to be [alleles](https://en.wikipedia.org/wiki/Allele) for this base position.

Single-nucleotide [polymorphisms](https://en.wikipedia.org/wiki/Polymorphism_%28biology%29) may fall within coding sequences of [genes](https://en.wikipedia.org/wiki/Gene), [non-coding regions of genes](https://en.wikipedia.org/wiki/Noncoding_DNA), or in the [intergenic regions](https://en.wikipedia.org/wiki/Intergenic_region) (regions between genes). SNPs within a coding sequence do not necessarily change the [amino acid](https://en.wikipedia.org/wiki/Amino_acid) sequence of the [protein](https://en.wikipedia.org/wiki/Protein) that is produced, due to [degeneracy of the genetic code](https://en.wikipedia.org/wiki/Genetic_code#Degeneracy).



SNPs in the coding region are of two types, synonymous and nonsynonymous SNPs. Synonymous SNPs do not affect the protein sequence while nonsynonymous SNPs change the amino acid sequence of protein. The nonsynonymous SNPs are of two types: [missense](https://en.wikipedia.org/wiki/Missense_mutation) and [nonsense](https://en.wikipedia.org/wiki/Nonsense_mutation).

SNPs that are not in protein-coding regions may still affect [gene splicing](https://en.wikipedia.org/wiki/Gene_splicing), [transcription factor](https://en.wikipedia.org/wiki/Transcription_factor) binding, [messenger RNA](https://en.wikipedia.org/wiki/Messenger_RNA) degradation, or the sequence of [non-coding RNA](https://en.wikipedia.org/wiki/Non-coding_RNA). Gene expression affected by this type of SNP is referred to as an eSNP (expression SNP) and may be upstream or downstream from the gene.



Types of SNPs

Non-coding SNPs: They can be 5’-UTR (Untranslated Regions), 3’-UTR, introns, intergenic regions and pseudogenes occurring in the genome. The regulation machinery of these noncoding SNPs includes splicing, transcriptional regulation and translational regulation.

Coding SNPs: These types of SNPs are also called as synonymous SNPs (due to third position variation). Other coding SNPs are replacement SNPs which cause change in amino acid, therefore categorized again in two types: 1) Functional SNPs (acceptable amino acid replacement) and 2) Non-functional SNPs (traits & diseases)

The frequency of their occurrence varies from about 1 in 1000 bases to 1 in 100-300 bases. SNPs can help scientists to locate genes that are associated with specific diseases. SNPs play a vital role in disease development when they occur within a gene or in a regulatory region near a gene.

Although SNPs are commonly occurring throughout the human genome and preferably do not have any side effects. Although scientists believe that SNPs may be proved a cutting edge technology to identify the genes responsible for several deadly diseases such as cancer, diabetes, vascular disease, and some forms of mental illness.

**Applications of SNPs**

 There are several applications of SNPs in disease diagnosis and other aspects as follows:

* Use of SNPs in disease diagnosis: The genome of every individual has unique SNP pattern made up of different genetic variations. Although, most of these SNPs are housekeeping and does not cause any type of disease, yet they may serve as biological markers for identification of a particular disease as they are commonly occurring near to a gene associated with a disease. Occasionally, SNPs may also be responsible for a disease cause due to genetic variation. Therefore, scientists are actively involved in identification of genetic make up and SNP pattern of individuals affected with any disease and prepare a database for further use as reference.
* Use of SNPs in public databases: SNPs are present throughout the genome, thus they serve as biological marker i.e. have identifiable physical location that can be easily tracked and used for constructing a chromosome map and can show a position of a gene. NCBI has prepared a huge public SNP database (dbSNP) for comparision of genome of several species. This database can be assessed by any research community to compare and identify the genome sequence of species.
* Use of SNPs in drug development: The SNPs can also have importance in action of a particular therapeutic agent. Currently, no method is present to predict the reaction of any individual against a medication i.e a treatment which is proven effective in one patient may not work in others. Even some patients may experience the side effects. As a futuristic tool, SNPs can be used to predict the possible behaviour of an individual for particular therapeutic agent by analyzing his SNP profile, may be referred as “personalized medicine”. An overview of application of SNPs in Human genome and personalized medicine

**Junk DNA - repetitive sequences**

**Repetitive DNA** Eukaryote and also human DNA contains large portion of noncoding sequences. As for the coding DNA, the noncoding DNA may be unique or in more identical or similar copies. DNA sequences with high copy numbers are then called repetitive sequences. If the copies of a sequence motif lie adjacent to each other in a block, or an array, we are speaking about tandem repeats, the repetitive sequences dispersed throughout the genome as single units flanked by unique sequence are interspersed repeats.

**The nature of interspersed repeats -  transposable element (TE or transposon) :**

is a [DNA sequence](https://en.wikipedia.org/wiki/DNA_sequence) that can change its position within a [genome](https://en.wikipedia.org/wiki/Genome), sometimes creating or reversing [mutations](https://en.wikipedia.org/wiki/Mutation) and altering the cell's [genome size](https://en.wikipedia.org/wiki/Genome_size). Transposition often results in duplication of the TE. [**Barbara McClintock**](https://en.wikipedia.org/wiki/Barbara_McClintock)**'s** discovery of these **jumping genes** earned her a [Nobel Prize](https://en.wikipedia.org/wiki/Nobel_Prize) in 1983. There are essentially two types of transposable DNA elements, or transposons: DNA transposons and retrotransposons.



**How a typical (ClassI) Retrotransposons functions?**

For jumping they require cellular RNA polymerases (II or III) by which they are transcribed into RNA, while the original DNA copy is maintained at the same location. The RNA copy is reverse-transcribed into DNA, and the DNA is inserted into the genome at a new location. Thus, these elements expand in number by a duplication (copy-and-paste) mechanism. the L1 retrotransposon, process of retrotransposition is prone to various mistakes عرضة لأخطاء مختلفة, so the new copies of a retrotransposon would be largely inactivated, because of truncation الاقتطاع or point mutation. Because most of the transposon copies are inactive, the further expansion of the retrotransposon family is governed by the few active full-length elements. However, even if all the active elements were lost later during evolution, the genome might be literarly over run with the fossil members of the sequence family. Retrotransposons can be further classified as autonomous and nonautonomous. Autonomous retrotransposons are coding for proteins necessary for their transposition, although they are also dependent on host RNA polymerases and DNA repair enzymes for successful jumping. Nonautonomous retrotransposons do not code for any protein and must hijackيخطف other transposon´s enzymes to be able of transposition.



**How a typical (classII) DNA transposon functions?**

 The core of the transposable element codes for an enzyme transposase. This enzyme binds to the ends of the element. The ends of the transposon are formed by inverted repeats, which can therefore exchange DNA strands and stabilize the stem-loop structure necessary for transposase action. Transposase then cuts the transposon out and ligates the resulting free chromosomal DNA ends. The free complex transposon-transposase binds to a specific sequence motif elsewhere in the genome, transposase cleaves the host DNA and ligates the transposon into the new place. Thus, the transposon moves by a cut-and-paste mechanism and the copy number remains stable.





**Transposon in disease**

TEs are [mutagens](https://en.wikipedia.org/wiki/Mutagen) and their movements are often the causes of genetic disease. They can damage the genome of their host cell in different ways:

* a transposon or a retrotransposon that inserts itself into a functional gene will most likely disable that gene;
* after a DNA transposon leaves a gene, the resulting gap will probably not be repaired correctly;
* multiple copies of the same sequence, such as [Alu sequences](https://en.wikipedia.org/wiki/Alu_sequence), can hinderيمنع precise [chromosomal](https://en.wikipedia.org/wiki/Chromosome) pairing during [mitosis](https://en.wikipedia.org/wiki/Mitosis) and [meiosis](https://en.wikipedia.org/wiki/Meiosis), resulting in unequal [crossovers](https://en.wikipedia.org/wiki/Chromosomal_crossover), one of the main reasons for chromosome duplication.

**Tandem repeats**

Tandem repeats are made of succesive identical or nearly identical (degenerate) repeat units. They vary in length of repeat unit as well as lenght of the whole repeat much, so every classification is not satisfying .

When between 10 and 60 nucleotides are repeated  are typically repeated 5-50 times, it is called a [minisatellite](https://en.wikipedia.org/wiki/Minisatellite). Those with fewer are (ranging in length from 2–5 [base pairs](https://en.wikipedia.org/wiki/Base_pairs)) are repeated, typically 5–50 times known as [microsatellites](https://en.wikipedia.org/wiki/Microsatellite_%28genetics%29) or [short tandem repeats](https://en.wikipedia.org/wiki/Short_tandem_repeat). they have a higher [mutation](https://en.wikipedia.org/wiki/Mutation) rate than other areas of DNA leading to high [genetic diversity](https://en.wikipedia.org/wiki/Genetic_diversity).



When exactly two nucleotides are repeated, it is called a *dinucleotide repeat* (for example: ACACACAC…). The [microsatellite instability](https://en.wikipedia.org/wiki/Microsatellite_instability) in [hereditary nonpolyposis colon cancer](https://en.wikipedia.org/wiki/Hereditary_nonpolyposis_colon_cancer) most commonly affects such regions.

When three nucleotides are repeated, it is called a *trinucleotide repeat* (for example: CAGCAGCAGCAG…), and abnormalities in such regions can give rise to [trinucleotide repeat disorders](https://en.wikipedia.org/wiki/Trinucleotide_repeat_disorders).

When the repeat unit copy number is variable in the population being considered, it is called a [variable number tandem repeat](https://en.wikipedia.org/wiki/Variable_number_tandem_repeat) (VNTR).  is a location in a [genome](https://en.wikipedia.org/wiki/Genome) where a short [nucleotide sequence](https://en.wikipedia.org/wiki/Nucleotide_sequence) is organized as a [tandem repeat](https://en.wikipedia.org/wiki/Tandem_repeat). These can be found on many [chromosomes](https://en.wikipedia.org/wiki/Chromosome), and often show [variations](https://en.wikipedia.org/wiki/Polymorphism_%28biology%29) in length between individuals. Each variant acts as an [inherited](https://en.wikipedia.org/wiki/Heredity) [allele](https://en.wikipedia.org/wiki/Allele), allowing them to be used for personal or parental identification. Their analysis is useful in [genetics](https://en.wikipedia.org/wiki/Genetics) and [biology](https://en.wikipedia.org/wiki/Biology) research, [forensics](https://en.wikipedia.org/wiki/Forensics), and [DNA fingerprinting](https://en.wikipedia.org/wiki/DNA_fingerprinting).



**Applications of Tandem repeats**

1. Tandem repeat describes a pattern that helps determine an individual's inherited traits.
2. Tandem repeats can be very useful in determining parentage. [Short tandem repeats](https://en.wikipedia.org/wiki/Short_tandem_repeat) are used for certain [genealogical DNA tests](https://en.wikipedia.org/wiki/Genealogical_DNA_test) (looks at a person's genome at specific locations for the purposes of determining ethnicityعرق and genealogical النسب relationships).
3. Polymorphic tandem repeats (alias VNTRs) are also present in microorganisms and can be used to trace the origin of an outbreak.
4. In the field of [Computer Science](https://en.wikipedia.org/wiki/Computer_Science), tandem repeats in strings (e.g., DNA sequences) can be efficiently detected using [suffix trees](https://en.wikipedia.org/wiki/Suffix_tree) or [suffix arrays](https://en.wikipedia.org/wiki/Suffix_array).