

## Single-nucleotide polymorphism

---

A **Single Nucleotide Polymorphism**, also known as **Simple Nucleotide Polymorphism**, (SNP, pronounced *snip*; plurals *snips*) is a DNA sequence variation occurring commonly within a population (e.g. 1%) in which a single nucleotide — A, T, C or G — in the genome (or other shared sequence) differs between members of a biological species or paired chromosomes. For example, two sequenced DNA fragments from different individuals, AAGCCTA to AAGCTTA, contain a difference in a single nucleotide. In this case we say that there are two *alleles*. Almost all common SNPs have only two alleles. The genomic distribution of SNPs is not homogenous; SNPs occur in non-coding regions more frequently than in coding regions or, in general, where natural selection is acting and 'fixing' the allele (eliminating other variants) of the SNP that constitutes the most favorable genetic adaptation.<sup>[1]</sup> Other factors, like genetic recombination and mutation rate, can also determine SNP density.<sup>[2]</sup>

SNP density can be predicted by the presence of microsatellites: AT microsatellites in particular are potent predictors of SNP density, with long (AT)(n) repeat tracts tending to be found in regions of significantly reduced SNP density and low GC content.<sup>[3]</sup>

Within a population, SNPs can be assigned a minor allele frequency — the lowest allele frequency at a locus that is observed in a particular population. This is simply the lesser of the two allele frequencies for single-nucleotide polymorphisms. There are variations between human populations, so a SNP allele that is common in one geographical or ethnic group may be much rarer in another.

These genetic variations between individuals (particularly in non-coding parts of the genome) are sometimes exploited in DNA fingerprinting, which is used in forensic science. Also, these genetic variations underlie differences in our susceptibility to disease. The severity of illness and the way our body responds to treatments are also manifestations of genetic variations. For example, a single base mutation in the APOE (apolipoprotein E) gene is associated with a higher risk for Alzheimer

## Types

---

Single-nucleotide polymorphisms may fall within coding sequences of genes, non-coding regions of genes, or in the intergenic regions (regions between genes). SNPs within a coding

sequence do not necessarily change the amino acid sequence of the protein that is produced, due to degeneracy of the genetic code.

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 and nonsense.

SNPs that are not in protein-coding regions may still affect gene splicing, transcription factor binding, messenger RNA degradation, or the sequence of 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.

### **Use and importance**

---

Variations in the DNA sequences of humans can affect how humans develop diseases and respond to pathogens, chemicals, drugs, vaccines, and other agents. SNPs are also critical for personalized medicine.<sup>[5]</sup> However, their greatest importance in biomedical research is for comparing regions of the genome between cohorts (such as with matched cohorts with and without a disease) in genome-wide association studies.

The study of SNPs is also important in crop and livestock breeding programs. See SNP genotyping for details on the various methods used to identify SNPs.

SNPs are usually biallelic and thus easily assayed.<sup>[6]</sup> A single SNP may cause a Mendelian disease. For complex diseases, SNPs do not usually function individually, rather, they work in coordination with other SNPs to manifest a disease condition as has been seen in Osteoporosis.

**On the other site, all types of SNPs can have observable phenotype or can result in disease:**

- SNPs in non-coding regions can manifest in higher risk of cancer<sup>[18]</sup>
- SNPs in coding regions:
  - synonymous substitutions by the definition do not trigger amino acid change in the protein, but still can affect its function e.g. seemingly silent mutation in the multidrug resistance gene 1 (MDR1), which codes for a cellular membrane pump that expels drugs

from the cell, can slow down translation to allow the peptide chain to fold into an unusual conformation causing the mutant pump be less functional<sup>[19]</sup>

- nonsynonymous substitutions:
  - missense - single change in the base results in change in amino acid of protein and its malfunction which leads to disease (e.g. c.1580G>T SNP in LMNA gene - position 1580 (nt) in the DNA sequence (CGT codon) causing the guanine to be replaced with the thymine, yielding CTT codon in the DNA sequence, results at the protein level in the replacement of the arginine by the leucine in the position 527,<sup>[20]</sup> at phenotype level this manifest with overlapping mandibuloacral dysplasia and progeria syndrome)
  - nonsense - point mutation in a sequence of DNA that results in a premature stop codon, or a *nonsense codon* in the transcribed mRNA, and in a truncated, incomplete, and usually nonfunctional protein product (e.g. Cystic fibrosis caused by the G542X mutation in the cystic fibrosis transmembrane conductance regulator gene).

## SNP analysis

---

Analytical methods to discover novel SNPs and detect known SNPs include:

- DNA sequencing,<sup>[32]</sup>
- capillary electrophoresis,<sup>[33]</sup>
- mass spectrometry,<sup>[34]</sup>
- single-strand conformation polymorphism (SSCP);<sup>[35]</sup>
- electrochemical analysis;
- denaturing HPLC and gel electrophoresis;
- restriction fragment length polymorphism;
- hybridization analysis;