**Course: Advanced Bioinformatics**

**Module title: Single Nucleotide Polymorphisms**

**Module no: 118**

A single nucleotide polymorphism or simple nucleotide polymorphism, often abbreviated to just SNP (pronounced snip; plural snips), is a variation in a single nucleotide which may occur at some specific position in the genome, where each variation is present to some appreciable degree within a population (e.g. >1%).

For example, at a specific base position in the human genome, it may be that in most individuals the base C appears there; but in a minority of individuals, the base A appears at that position instead. There is an SNP at this specific base position, and the two possible nucleotide variations - C or A - are said to be alleles for this base position. Although in this example and most SNPs so far discovered there are only two different alleles, there are also triallelic SNPs in which three different base variations may coexist within a population.

SNPs underlie differences in our susceptibility to disease; a wide range of human diseases, e.g. sickle-cell anemia, β-thalassemia and cystic fibrosis result from SNPs. 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's disease.

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.

Frequency

Within a genome: 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. Other factors, like genetic recombination and mutation rate, can also determine SNP density.

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.

Within a population: 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. 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.