**Introduction to Bioinformatics**

**Kishan Sarpangala**

**Discuss briefly the implications of mismatches due to polymorphisms and sequencing errors**

**What is polymorphism?**

The presence of genetic variation within a population, upon which natural selection can operate.

Due to polymorphism there could be mismatch on a subset of human population.

* False positive can happen on a nucleotide **polymorphism** while reading/sequencing.
* Other issue is No match or alignment wouldn’t take place.

**How does sequencing error effect in genomics?**

Even though the sequencing accuracy for each individual nucleotide is very high, the very large number of nucleotides in the genome means that if an individual genome is only sequenced once, **there will be a significant number of sequencing errors**. Furthermore, many positions in a genome contain rare [single-nucleotide polymorphisms](https://en.wikipedia.org/wiki/Single-nucleotide_polymorphism) (SNPs). Hence to distinguish between sequencing errors and true SNPs, it is necessary to increase the sequencing accuracy even further by sequencing individual genomes a large number of times.

### Ultra-deep sequencing and Transcriptome sequencing

Is used for higher coverage in sequencing

When we look at software point of view it would be very important that the program enables the user to infer error profiles based on sequencing data for a given genome. Erroneous base calls should be identified based on the alignment of the reads against the known reference of the genome. The software must be able to identify mismatches, insertions and deletions for any sequenced data set. Else we will be wasting lot of precious research time.