**Chapter 2**

Mutations can occur in cells at all stages of life and in all tissues and clonality and classification is dependent on tissue-type, size of the genetic modification and time of the mutation. Germline mutations are mutations the genetic polymorphisms that children inherits from their parents. Child might have mutations that are private from their parents as a result of de novo mutations occuring during cell division and mutagenesis during meiotic recombniation (discussed in Chatper 5, reference). Somatic mutations are mutations that occur post-fertilisation. If somatic mutation occurs during first-cell division of embryonic development, for example, somatic mutation will be present on all the daugther cells originating from the embroynic cell with the somatic mutation. Somatic mutations can be differently classified depending on ther size and their orientation. Most somatic mutations are benign, but some somatic mutations confer proliferative advantage to the cell and these somatic mutations are referred to as driver mutations. Somatic muations can be substitution, small insertions and deletions (<50bp) or large changes to genomes (>50bp), also known as structural variations. The advent of next-generation sequencing has enabled us to detect more complex forms of genomic rearrangemetns such as chrmothripsis, chromoplexy, break-induced replication, fork stalling and template switching, etc and the elucidation of DNA damage and repair pathways that generate these complex rearrangements. Somatic mutation detection, hence, is often the first-step towards characterising the cancer genome. Thanks to decrease in cost for whole-genome sequencing, thousands of cancer genomes have been sequenced, tissue-specific driver mutations and somatic mutational processes have been uncovered. However, due to the technical limitations of the NGS platform limit the length, clonality, diversity and resolution of the somatic mutations that can be detected. The base quality score, measuring the probability that the base read out is a sequnencing error and not a true representasion of the template molecule, ranges from 0 to 40, 1 error per 10,000 bp, for Illumina short reads and as a result, somatic mutation caller such as Mutect [reference], Strelka2 [reference] requires matched tumour and normal sequencing to distinguish germline mutations from somatic mutations and cannot distinguish low allelic fraction somatic mutations from sequencing errors. In addition, repetitive sequences account for 50\% of the human genome and repeat length is often greater than the short read length. Moreover, there isn't one type of repeat, multiple types of repeats are present in the human genome (tandam repeat expansions, retrotransposons like ALUs and SINEs, non-repetitive unique sequence copies such as segmental duplications, alpha-satellites in centromeric regions and centromeric repetitive sequences) and these repetitive structures are responsible for somatic mutagenesis both small and big. If the repeat length is greater than the read length, read aligners that determines the position of read relative to the reference genome cannot ascertain the position of the read relative to the reference genome and these reads without fixed locations cannot be used to call somatic mutations. To determine the location of the read relative to the reference genome with high-confidence, the repetitive sequences within the read must be flanked by unique sequences not exisiting elsewhere in the reference genome. In addition, somation detection isn't a solved problem. Different somatic mutation callers with different strategies for somatic mutation calling have different sensitivity and specificity and often the consensus of different somatic mutation caller provides the most accurate somatic mutation calls. In addition, if tissue in question doesn't have sufficent DNA and requires PCR amplification, somatic mutation detection might also be confounded with PCR amplifiaction bias.