Scriber: Huen Long Yin
SID: 1155159568

Lecture 16: Cancer Genomics overview

Friday, 27th October 2023

Lecture Outline:

- Cancer genomics overview
- Genome: variant calling

1. Cancer genomics overview

1.1 What is cancer?

- A disease in which some of the body's cells grow **uncontrollably** and **spread** to other parts of the body

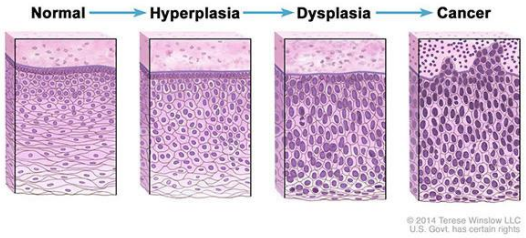


Figure 1 Formation of cancer

1.2 Why do we want to study cancer?

- More than half a million people in the US died from cancer in 2021
- Cancer remains a leading cause of death globally

1.3 How do we study cancer?

- Cancer is usually believed to be a **genomic** disease
    - **Study method: genomics/ multi-omics**
    - Examples: genome, epigenome, transcriptome, proteome, metabolome

1.4 Data analytics for cancer genomics

- Genome analysis: variant calling, genome association study
- Epigenome analysis: identification of gene, peak calling, differential peak calling
- RNA-seq: DEG, gene fusion

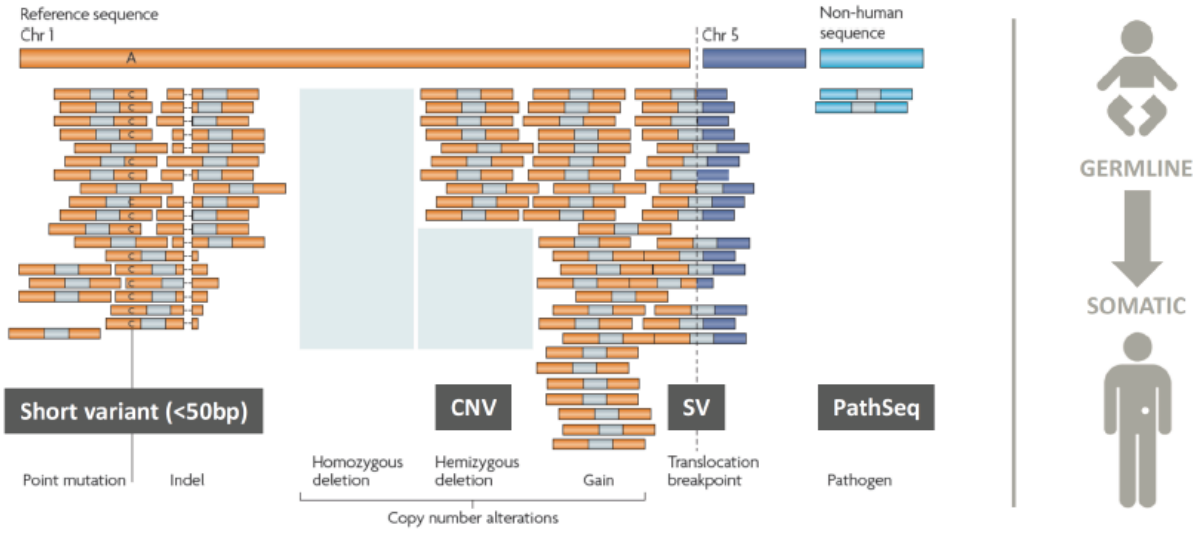
Scriber: Huen Long Yin
SID: 1155159568

2 Genome: variant calling

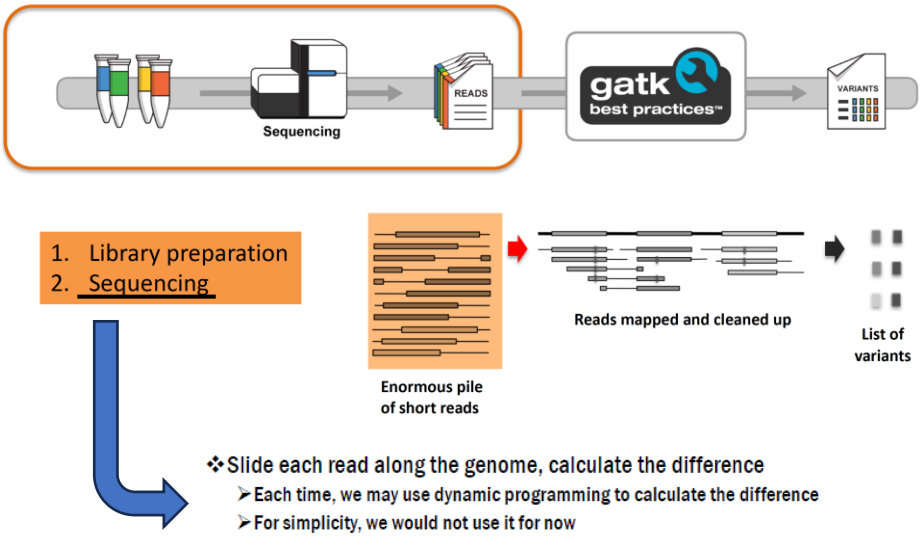
2.1 Why do we care about gene variants?

- 3.2 billion sites in the human genome
  - Any 2 humans share **99.5%** DNA → Can easily **describe a genome** with relation to a reference
- Gene differences → differences in disease risk & response to treatment
  - Cancer can be considered as genetic variants at **multiple levels**
- Genetic variation is used to find genes and variants that **contribute to disease**

2.2 Different types of gene variants



2.3 How to discover the genetic variants?



Scriber: Huen Long Yin
SID: 1155159568

2.4 How to distinguish **actual variation** (real change) and **errors**(artifacts) in the analysis?

Types of errors that could occur at different steps of the analysis:

1. PCR artifacts (amplification of errors)
2. Sequencing (errors in base calling)
3. Alignment (misalignment, mis-gapped alignments)
4. Variant calling (low depth of coverage, few samples)
5. Genotyping (poor annotation)

View of probable variants in a genome browser: