BMEG 3105 Fall 2022

Data analytics for personalized genomics and precision medicine:

Lec-16: Cancer Genomics Overview

Cancer?

- ✧ Some of the body cells grow uncontrollably
- ✧ Can spread to other parts of body
- ✧ Caused a lot of death
- ✧ Usually believed to be genomic disease
- ➢ Genomics/multi-omics methods【Genome/Epigenome/Transcriptome/Proteome/Metabolome】

Data analytics for cancer genomics

- ⌑ Genome: variant calling, genome association study
- ⌑ Epigenome: what is it, peak calling, differential peak calling
- ⌑ RNA-seq: DEG, gene fusion

❖ Variant 【*Find out more if interested: http://software.broadinstitute.org/gatk/*】

- ✧ Short variant (< 50 bp)
  - ○ Point mutation: one base is changed
  - ○ InDel (one base Inserted/Deleted)
- ✧ CNV (Copy Number Alterations)
  - ○ Homozygous deletion
  - ○ Hemizygous deletion
  - ○ Gain
- ✧ SV (Structural Variant)
- ✧ PathSeq (from Pathogen)

Heritability?
- ✓ Germline (on sperm/egg)
- × Somatic (on other body cells)

How to discover genetic variants?
Do sequence mapping (Lec-4)\

Actual variation (real change) vs. error (artifact)
Errors can creep in on various levels:
- ⌑ PCR artifacts (amplification of errors)
- ⌑ Sequencing (errors in base calling)
- ⌑ Alignment (misalignment, mis-gapped alignments)
- ⌑ Variant calling (low depth of coverage, few samples)
- ⌑ Genotyping (poor annotation)