**Quiz 4**

1. Multi-omics 데이터의 종류에 대해 최소 3가지 이상 나열하고, 각각의 데이터가 포함하고 있는 정보를 설명하세요.

* Copy number variation: DNA sequencing 또는 array를 통해 획득. 유전자 copy수의 gain 또는 loss에 대한 정보를 포함
* Gene differential expression: RNA sequencing을 통해 획득. 유전자 발현량의 차이 대한 정보를 포함
* Methylation: Epigenomics 분석으로 획득. 유전자 발현 억제에 대한 정보 획득.
* ncRNA expression: 단백질로 translation되지 않는 non-coding RNA에 대한 발현량 정보를 포함. 유전자의 발현을 조절에 대한 정보를 얻을 수 있음

1. Hardy-Weinberg equilibrium (HWE)과 linkage disequilibrium (LD)의 개념을 설명하고, LD가 발생할 수 있는 요인에 대해 나열하세요.

* Hardy-Weinberg equilibrium (HWE): Hardy와 Weinberg가 만든 수학적 모델.  
  Selection, genetic drift, mutation 등의 영향이 없으면 allele frequency는 안정되어 있음.  
  Bi-allelic variant, random mating, infinite population, no migration, no selection이 해당하지 않는다면 HWE가 성립하지 않는다.
* Linkage disequilibrium (LD): allele들이 단단하게 연결되어 있어서 random하지 않게 association되어 있는 경우 LD라고 한다.
* LD 발생 요인: occurring new mutation, population admixture, natural selection, reduced recombination

1. RNA-seq 데이터의 분석 파이프라인에 대해 서술하세요 (raw data인 fastq 파일부터).
   * Sequencing을 통해 얻은 fastq files을 genome 또는 Transcriptome에 mapping
   * Gene 또는 transcriptome에 mapping된 reads (sam or bam files)를 counting
   * Sample 조건 또는 실험과정에서 발생하는 변인들을 제거해주는 normalization을 실시
   * Differential expression, variant calling, annotation, novel transcript discovery, RNA editing 등을 분석
2. Virtual screening의 종류를 나열하고 각각의 방법에 대해 간단히 서술하세요.

|  |  |  |
| --- | --- | --- |
| 3D Structure of Target | Ligand-Based Methods | Similarity searching |
| Pharmacophore mapping |
| Machine learning methods |
| Structure-Based Methods | Protein ligand docking |

* + Similarity searching: 구조는 잘 모르는 protein target이 있을 때, 해당 target의 activity에 영향을 주는 물질들이 알려져 있다면, 그 물질과 유사한 물질을 찾는 방법. Drug-likeness(rule of five: RO5) 안에서 searching. 이때 분자구조를 chemical fingerprint로 변환하여 tanimoto coefficient를 이용하여 유사도를 계산.
  + Pharmacophore mapping: 특정 target에 대해 지금까지 activity가 있다고 알려진 물질들의 공통적 특징(분자 구조적)들을 찾아서 screening
  + Machine learning methods: 지금까지 축적된 data에서 특정 target에 대한 물질들의 activity와 inactivity 정보를 가지고서 학습을 실시. QSAR이 대표적. 분자 구조를 숫자형태로 변환하고, 해당 물질의activity 정도를 수치화해서 학습.
  + Protein ligand docking: target protein의 구조를 알 때, target에 ligand를 가상으로 binding 시켜보는 것으로, docking 연산에는 pose prediction과 binding affinity를 이용하여 계산하고, 이를 이용해서 특정 target과 ligand의 activity를 예측해 볼 수 있다.