PLPTH613, Homework 5

Due day: May 9th

The purpose of this homework is to strengthen your understanding for QTL, GWAS, genome assembly.

**Problem 1.**

Describe the procedure of the QTL permutation for identifying the significant LOD threshold. Briefly discuss what is the problem if we do not control population structure in a GWAS analysis.

**Problem 2.**

Given the following reads that contain no sequencing errors, perform an OLC assembly using the parameters of at least 5 bp overlap and 100% identity. Draw the overlap graph, and report the resulting assembled sequence.

1. GCATGCA

2. TGCATGC

3. CGTGCAT

4. CATGCAG

5. GTGCATG

**Problem 3.**

Given the following reads that contain no sequencing errors, perform a de Bruijn graph assembly using k-mer = 3 and 6 separately. List the k-mers, draw the de Bruijn graph, and report the resulting assembled sequence.

1. GCATGCA

2. TGCATGC

3. CGTGCAT

4. CATGCAG

5. GTGCATG

**Problem 4.**

1) What is the problem if we select the BWA aligner instead of the GSNAP aligner for aligning RNA-Seq reads to a reference genome sequence?

2) Describe the reason for performing multiple testing control in RNA-Seq analysis. How to select significant genes based on the list of q-values?