Bios 560R High-throughput data analysis using R and Bioconductor

Homework 5

Due on **Nov 29th, Tuesday before class at 1pm**.

I. Read Wikipedia pages for “DNA sequencing”,

II. Short answer questions, 10 points each. Be creative in answering the questions.

1. What is DNA sequencing? Why do we want to do DNA sequencing?
2. Compared to traditional sequencing method (Sanger sequencing), what are the pros and cons of second-generation sequencing?
3. Briefly describe the workflow of second generation sequencing data analysis.
4. Suppose after one run, the sequencing machine generated 1 million sequence reads, each of 50 base pairs long. What will be the dimension of the raw intensity data?
5. What’s the difference between sequence alignment and assembly?