BIOS 497/897-1

HW4: short read mapping

Due on Feb. 12th, 11:59PM

1. Two data files are included in this package. One, refgenome.fa, is the reference genome, and the other, 50\_reads.fastq, is the fastq file for 50 reads.

2. Downloading Bowtie

<http://bowtie-bio.sourceforge.net/index.shtml>

3. Installing Bowtie

There is no a real installation step for Bowtiw. After unzipping the package, you may find the bowtie executable file, and run it directly.

Tutorial

<http://bowtie-bio.sourceforge.net/tutorial.shtml>

4. Using Bowtie.

For MS Windows, you need to open a DOS window (like: start -> run -> cmd). For Linux or Mac OS, you need to open a terminal window.

For a simple way, you may copy all sequence files to the directory that has the bowtie executable files.

Then, you may use bowtie with the following commands.

step 1: indexing

> bowtie-build -f refgenome.fa refgenome.index

step 2: mapping

> bowtie -v 2 -S --best --strata -k 10 refgenome.index -q 50\_reads. fastq i.sam

5. Understanding the output

How many mismatches are allowed for mapping if using the above command line? Opening the output alignment file, for example “i.sam” in the above example. Looking for uniquely aligned reads. How many uniquely aligned reads can you find? You may manually count them or make a program to calculate. You can get extra credits if you make a program. If you use perl, “hash” is suitable to this task.

To submit your homework, please show the command lines you used to run bowtie (it is better to have a screen shot), the output file, and answers to questions. If you make a program, please submit your source code file.