1. A computer algorithm is

- ⦿ A precise specification of all the steps needed to compute a solution to a problem
- ◯ A protocol for transmitting data over a network
- ◯ A program written in a computer language such as C++, Python, or Perl
- ◯ A description of the hardware and software capabilities of a computer

2. You can make a program more efficient by

- ⦿ Re-designing the algorithm to require fewer steps
- ◯ Recompiling it on a different computer
- ◯ Running it on the Amazon cloud
- ◯ Replacing your hard drives with faster solid state drives

3. DNA sequences can be represented efficiently using

- ⦿ Two bits for each of the 4 possible bases
- ◯ The SAM alignment format
- ◯ The compressed BAM file format
- ◯ One byte for each of the 4 possible bases

4. A programming language is

- ● A formal language used to instruct computers what to do
- ○ A web-based environment for building computer programs
- ○ Anything written in Python
- ○ A method for translating between languages such as English and French

5. Software engineering involves

- ○ Documenting computer code so that others can understand it
- ○ Updating code so that it remains compatible with other software systems.
- ○ Testing programs on a wide range of examples to see if they perform as expected
- ● All of these options

6. Bowtie, TopHat, and Cufflinks are

- ● Programs for analysis of RNA-seq (transcriptome) data sets
- ○ A web-based system for browsing genome data
- ○ Elements of men's formal evening wear
- ○ Programs for determining the function of a gene

7. Sequence alignment refers to

○◉ Lining up two DNA sequences so that positions with the same base match one another

○ Making sure that all the DNA sequences in a file use the same format

○ Shuffling the positions in a sequence to randomize them

○ Determining the amino acid sequence produced by translating a DNA sequence

8. A software pipeline for RNA sequence (RNA-seq) analysis will

◉ Process large raw sequence files into a summary table showing which genes were present

○ Create a database in which one can efficiently store and retrieve results

○ Automatically update all software to the latest version

○ Compare cases to controls and determine which genes were responsible for any differences in phenotype

9. Which of the following is not a computer operating system?

○ Mac OS X

○ Microsoft Windows

○ RedHat Linux

◉ Google Drive

10. A data set large enough to overwhelm the main memory of a computer would

   ○ Contain more than 1000 genomes' worth of data

   ○ Use uncompressed files rather than compressed ones

   ● Be larger than the available RAM

   ○ Be at least 100 gigabytes in size