## CBB752 Final Quiz, Spring 2010

Name: \_\_\_\_\_

200 points in total

- 1. Explain the difference between local and global optimization [8 points].
- 2. What is the difference between a deterministic and stochastic model? Which term applies to Ordinary Differential Equation (ODE) models? [8 points]
- 3. In the context of mathematical modeling, what is an F test used for? [12 points]
- 4. Assume that B and M are two different cell types (e.g., Naïve B cells (B) and memory B cells (M)). Write a brief description of the potential meaning for each parameter in the model: [20 points]

$$\frac{dB}{dt} = s + pB - cB - tB$$
$$\frac{dM}{dt} = tB - cM$$

5. Describe how 454 sequencing (pyrosequencing) and Illumina sequencing (sequencingby-synthesis) work. Describe the main type of error that occurs during 454 sequencing and why it occurs? Why is the read length produced by the Illumina platform limited? [20 points]

6. What are the advantages of single-molecule based sequencing compared to amplification based approaches? [10 points]

7. What are the two main confounding factors in DNA sequence assembly? In this context, describe the meaning of the N50 statistic. [10 points]

- 8. The majority of reads in a ChIP-Seq experiment are \_\_\_\_\_. [5 points]
- 9. When scoring ChIP-Seq data, what is the purpose of the peak shift? What statistical distribution is typically used to assess significance of a ChIP-Seq peak? [15 points]

10. After running a ChIP-Seq experiment targeting factor X, you find that only 30% of the peaks contain the known motif for factor X. Assuming no experimental error, explain using *diagrams and text* the possible reasons why the motif may be underrepresented. [20 points]

11. In detail using text and diagrams, explain how an index-based short-read mapping algorithm works. Why are spaced-seeds necessary when using this approach? Make sure to show the spaced seeds approach in your diagram. [20 points]

12. What is the main drawback of microarrays compared to RNA-Seq? Name three new applications using RNA-Seq that were not possible with tiling arrays. [10 points]

- 13. Rank the following program in terms of sensitivity (1 = most sensitive, 6 = least sensitive) for finding sequence alignments: [12 points]
  - Bowtie
    BLAST
    Smith-Waterman
    HMM
    PSI-BLAST
    Sequence Profiles
- 14. What are the advantages of using a Biplot to represent a data matrix? [10 points]

15. Explain how PCA works and why it is relevant to biological data sets? What is the interpretation of the principal components and what are the loadings? Explain the role of PCA in eigenfaces analysis presented in class. [20 points]