CMM262/BGGN237 Midterm Exam 2018

Part 1: Short answer questions (please answer 6 of the 8 questions using 3 to 6 sentences per question).

Email questions and final answers to Ryan Marina: [rmarina@ucsd.edu](mailto:rmarina@ucsd.edu)

1. Describe the differences between single-end and paired-end sequencing. When might we opt to perform a paired-end experiment over a single-end, and vice-versa? What is an index sequence and why do we include it our library?
2. In sequencing experiments, what do we mean by cluster density and library complexity? How might these factors impact the quality of a sequencing experiment? What measures can be taken to alleviate these problems?
3. Before we can begin our RNA sequencing library preparation protocol, we need to perform an additional stage of RNA selection. Why is this necessary? What different methods can be used for RNA selection, and how might the specific type of technique influence the profile of your final library?
4. As we covered the RNA seq processing pipeline, we encountered several different file formats, including fastq and SAM/BAM files. Describe these file types and what information can be found each. In your answer, please briefly describe what a CIGAR string is, as well as what it represents with respect to alignment information.
5. In RNA sequencing experiments, what are PCR duplicates? When might PCR duplicates be problematic and why? And what additional steps could you factor into your experimental design to account for PCR duplicates?
6. Over the first few modules, you were introduced to the concept of a counts matrix in quantifying gene expression. Explain what a counts matrix is and describe its structure (i.e. what information does each column/row convey?). How does this information differ between a single-cell RNA sequencing experiment and a bulk sequencing experiment?
7. What are the advantages of single-cell RNA sequencing over bulk RNA sequencing? What are some limitations of single-cell sequencing? Describe an example in which it might be beneficial to perform singe-cell RNA seq over bulk RNA sequencing. Why?
8. Why is it necessary to perform read normalization? With respect to individual genes, what qualities need to be accounted for when performing normalization? How would this differ between a bulk RNA sequencing experiment and a 3’-end based single-cell RNA seq experiment, and why?