Bioinformatics and Genomics (BZ360) Fall 2018

Study guide for exam 3. The dates reference the exercises and slides posted on the schedule.

Non-coding RNAs and other genomic features: 10/30/18

1. What is a gene?
2. What proportion of RNA in a cell is mRNA?
3. What are transposable elements and what are the two general types?
4. What genomic features are commonly associated with transposons?
5. What are pseudogenes?
6. What are the three classes of small RNAs and what distinguishes them?
7. What types of genes do each class of small RNAs commonly regulate?
8. What two features unify the different classes of small RNAs?
9. What are tRNAs and how long are they?
10. What are rRNAs?
11. What proportion of RNA in a cell is rRNA?
12. What are long non-coding RNAs (lncRNAs) and what distinguishes them from mRNAs?
13. What is Xist?
14. What three steps are essential for small RNA sequencing library preparation (answer: adapter ligation to either ends of the small RNAs, reverse transcription, and PCR amplification)?

Small RNA-seq: 11/1-8/18 (see slides and exercises 10-12)

1. What software did we use to trim adapter sequences from small RNAs?
2. What software did we use to map small RNA reads to a reference genome?
3. What software did we use to identify the numbers of reads for specific small RNAs?

RNA-seq: 11/15-29/18 (see slides and exercises 13-16)

1. What is RNA-seq and what can it tell us about gene expression?
2. What are 5 general steps in RNA-seq library preparation?
3. What are the general data analysis steps in RNA-seq and related high-throughput sequencing applications (at least 4 major ones starting with quality control)?
4. What files are typically needed for RNA-seq data analysis?
5. What is gff format and what information is contained in a gff file and in each column and row of a file?

6. What file format are genome sequences commonly stored in?
7. What file format is mapping data stored in (such as the output from bowtie or tophat)?
8. What is a distinct challenge in aligning RNA-seq reads compared to aligning genome sequencing reads?
9. What software did we use to trim adapters and low quality bases?
10. What software did we use to align RNA-seq reads?
11. What software does TopHat use to do the actual alignments?
12. What software did we use to compare gene expression between brain and reference samples?
13. What software did we use to visualize the RNA-seq data?
14. What is R studio?
15. What is fpkm?
16. What are regular expressions?
17. What is the regular expression for a wild card character?
18. What is the regular expression *?
19. What is grep and what is the basic command structure?
20. What information is returned by grep?