CSE 185 Quiz 3 Topics List- Spring 2019
Quiz 3 will cover the topics listed below. Expect 1-2 questions from each category

ChIP-sequencing+epigenomics
Know at a high level what are: transcription factor vs. histone modification vs. whole cell extract,
Be familiar with IGV visualization of peaks, coverage profiles, and genes

Motifs and enrichment analysis
PFM (position frequency matrix): PFM[i,j] gives number of times nucleotide i seen at position j
PWM (position weight matrix): PWM[i,j] = log2(p[i,j]/pi[i]), where p[i] gives background nucleotide frequencies
How to score motifs using a PWM
Limitations of PWMs:

- Indels won't match
- Don't model dependencies between bases

Enrichment analysis:

- Contingency tables
- Interpreting output (p-values vs. odds ratio)

Single-cell RNA-sequencing design and analysis
Experimental design:

- Barcode vs. UMI vs. index:
    - Barcode: unique per cell
    - UMI: unique per molecule
    - Index: unique per sample
- For a given length L, there are 4^L possible barcodes (or UMIs)
- Beads and cells get loaded to droplets at a Poisson rate. E.g. if the mean number of beads per cell is mu, the probability that a droplet gets j beads is e^-j*mu^j/j!. Similar for number of cells per droplet.
- Capture rate vs. duplicate rate (see prelab 1)
- Barcode diversity = number of cells / number of barcodes

UNIX commands and command-line tools
Know what the following bioinformatics tools are used for: BWA, Homer, FIMO, bedtools, IGV
Know what types of data are stored in the following file formats: FASTA, FASTQ, SAM/BAM, VCF, BED

Know basics of UNIX commands: cat, head, tail, cut, grep, ls, cd, awk, sed, datamash
Know how to write the output of a command to a file:

- "My command…" > file.txt # writes output of a command to a file
- "My command…" >> file.txt # appends to an existing file

Know how to pipe output of one command as input to the next command