Administrivia:

- Room may change, look out in email

Project Ideas:

- All open ended, underspecified
- First three ideas are associated with RNAseq data
    - Explore what biases there are and information about them

1. Tools for visualizing bias

- Generate some figures/summary statistics
- Invent other statistics and wrap it up in a convenient user interface to help biologists/those who are analyzing the data
- This project handles HCI issues
- Interpret signatures in the data to learn more about the causes of these biases

2. Bias Distorts Allele Specific Expression Analysis?

- Typically both copies of alleles are expressed, but there are situations in which only one of alleles is expressed (a few dozen examples known in humans and mice)
- What is the implication of technical biases for this kind of analysis?

3. Impact of bias in other RNAseq use cases

- RNA structure prediction
    - Less than perfect computational methods
    - How can this be more accurate using RNAseq data?
    - Bias may be relevant if you're seeing overrepresentation/missing pieces that weren't there
- Ribosome speed and functional implications (footprinting)
    - Take RNA out of cells with ribosomes attached, freeze it
    - Seeing overrepresentation from one part of RNA means that there could be bias or that the ribosome spends a lot of time there
    - What are the implications of biases?
    - Need to understand the protocol and the impact of bias
- Bias correction models need some sort of a background model - neutral data so that we can compare it to the biased data

4. Improved crossover detection

- Background:
    - Homozygous position: mom and dad's positions agree
    - Heterozygous position: they disagree
    - How do you find heterozygous sites?
        - Using DNAseq data
        - They try to get 30x coverage of every site
    - How do we know which order the heterozygous positions appear in?
        - Phasing problem: in which phase do they appear?
        - Look at the data to see what's consistently happening
    - Crossover problem:

    - Closely related individuals, genetic recombination may happen
    - Representing two different chromosomes
    - Genetic recombination is relevant in reproduction (diploid to haploid)
    - Important purpose: it shuffles the genes around
- Project: improved crossover detection
    - Is crossover distinguishable from a phasing error?
    - Is there a way to incorporate other information (phasing calls) to corroborate?
    - Goal: build a tool to find maximum likelihood estimate of number of crossovers, based on simple models of crossover/error
        - Use calculation from slides
    - Maximum Likelihood Estimation Problem sort of
    - Apply this to data from the project that Ruzzo was talking about (clone organisms)
        - “Trios”: look at crossover information in systems between mom/dad/child