Molecular and Genetic Epidemiology
Assignment 3
Due by noon 1/27 by email to Kirsten Kangelaris, kkangelaris@medicine.ucsf.edu

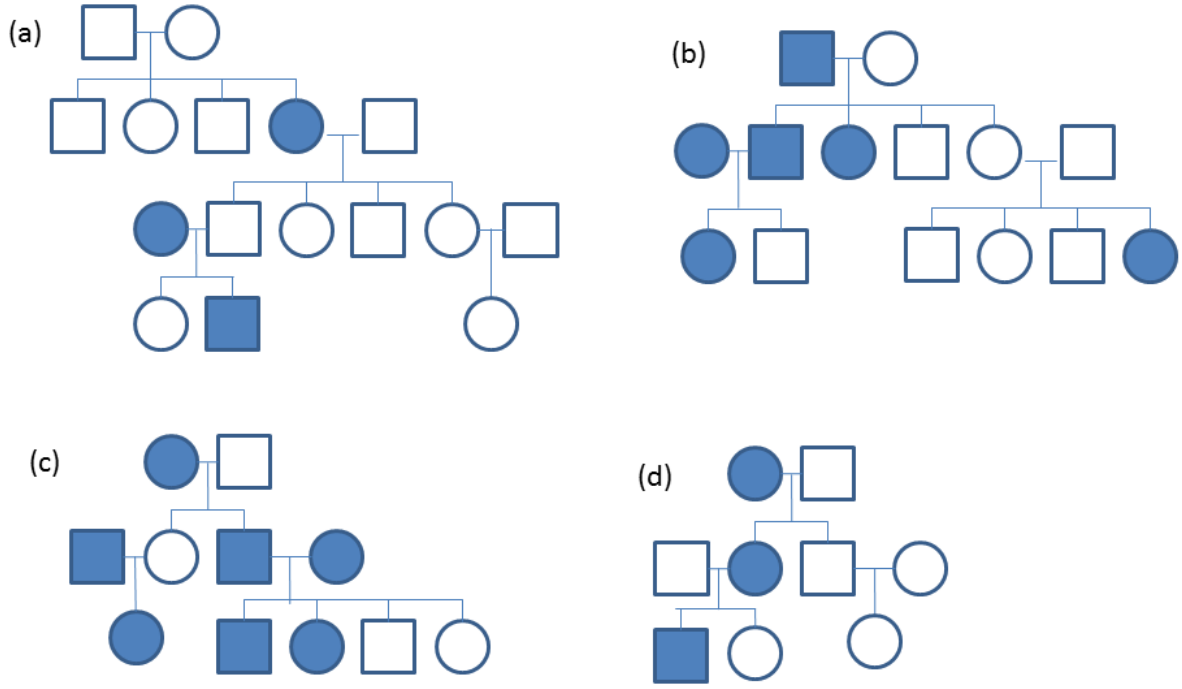
A. Textbook

- Reading Austin chapter 4

B. Problems

1. Suppose we are considering a Mendelian fully penetrant autosomal trait. If two recessive parents have the disorder, than what, if anything, can you say about their children? What if the trait were dominant? Now, suppose a child has a recessive trait, what, if anything, can you say about their parents? What if it were dominant?

2. What can you say about the following families – are they dominant (assuming fully penetrant disease), recessive (assuming fully penetrant disease), neither dominant nor recessive (assuming fully penetrant disease), or impossible to tell?



3. What are some of the differences between a microarray and sequencing that we have talked about?

C. Hands-on Analysis

We will be using data on two MTHFR SNPs (C677T, A1298C) for exercises in the remainder of the quarter. These are from a family-based (brothers) study of prostate cancer.

1. Walkthrough of loading the dataset into STATA (no answers required for this problem #1), as follows:

a. Download the STATA dataset “MTHFR.dta” from the course website.
(Note: you can also download an excel file for use with other programs)

b. Open this file (on completion of downloading, or by double clicking on it), to launch STATA. Or start STATA and the command

. use "C:\MTHFR.dta", clear

c. Open a Log in STATA by going to **File, Log, Begin**. Give the log a name (suggest using the “.log” format).

d. Look at the data with the following STATA commands:

. describe
. browse

As you will see, the data include information on genotypes defined by the two MTHFR SNPs, disease status, ethnicity, cancer aggressiveness, and what type of family the men are from. With regard to the latter, D=discordant means that the family includes at least one case and one unaffected brother (control). C=concordant means that the family has at least two affected brothers. C/D means that the family has at least two affected brothers, and one unaffected brother.

2. Calculate some allele frequencies, as follows:

a. Determine the genotype frequencies for these SNPs using the tabulate command in STATA.

tabulate MTHFR_C677T
tabulate MTHFR_A1298C

b. Repeat a, but stratified by ethnic group. Do the genotype frequencies differ by ethnic group?

tabulate MTHFR_C677T if Ethnicity=="Caucasian"
tabulate MTHFR_C677T if Ethnicity=="African-American"
tabulate MTHFR_A1298C if Ethnicity=="Caucasian"
tabulate MTHFR_A1298C if Ethnicity=="African-American"