Molecular and Genetic Epidemiology
Assignment 4
Due by noon 2/3 by email to Kirsten Kangelaris, kkangelaris@medicine.ucsf.edu

A. Textbook

- Reading Austin chapter 5

B. Project

We will continue to explore the potential genetic basis of "your" disease of interest (defined in Assignment 1). For this assignment, search PubMed for your disease and the terms "linkage analysis," "FBAT","TDT", and "haplotype" (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi).

How many, if any, articles are on PubMed? If any, briefly glance at some of the abstracts and write a couple of sentences summarizing what has been done.

Note the number of articles on PubMed, adding to your table from homework 2. Briefly glancing at some of the abstracts, give a qualitative assessment of whether or not there is positive or negative support for a genetic basis to the disease.

Search Terms	# of Articles on Medline	Positive Support	Negative Support
"Genetic Epidemiology"			
"Familial Aggregation"			
"Twins"			
"Heritability"			
"Segregation"			
"Linkage analysis"			
"FBAT(Family-based association test)/TDT(Transmission Disequilibrium Test)"			
"Haplotype"			

C. Problems

1. Linkage analysis. Suppose that we have a total of 50 recombinants, and 50 non-recombinants. What is the recombination fraction? Do you expect that this marker is linked to the trait?

2. Suppose we have a dichotomous trait in trios. Suppose that we have the following table of the affected offspring

		Non-transmitted parental allele	
		A	a
Transmitted parental allele	A	3	70
	a	30	700

Conduct the TDT test at this locus. Is the locus associated with the trait?

D. Hands-on Analysis

1. We will now undertake a family-based association analysis of the MTHFR SNP data that only includes the brothers discordant for prostate cancer (i.e., cases and controls). This is the file "MTHFRcase_control.dta".

a. Open this file to launch STATA. Or start STATA and type the command

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

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

Since the cases and controls are matched on being in the same family we should undertake a matched analysis. This can be done via conditional logistic regression (matching on the group family) first for all prostate cancers using the command
. clogit cap mthfr_c677t_new , group(familyid)
Is there a significant association; i.e., is the coefficient for the genotype significant?

We can also look at non-aggressive, and then aggressive prostate cancer, using the following two commands:
. clogit cap mthfr_c677t_new if aggressiveness==0, group(familyid)
. clogit cap mthfr_c677t_new if aggressiveness==1, group(familyid)
Does the association appear to be modified by aggressiveness?