Bioinformatics Workshop Exercise Handout #5

Day 3, AM: PLINK
Required Dataset: hapmap1 folder
Required Software: plink v1.9

Plink Exercise:
1. Check PLINK is installed on the system.
2. Download hapmap1.zip data and unzip to a new folder (HapMap).
3. Open and observe each file.
4. Identify how many individuals and SNPs are in the initial dataset?

Output File Exercise:
1. List the files in the current working directory.
2. Read the .bim file.
3. Read the .fam file.

Missing Data Exercise:
Determine the amount of missing data.
1. Produce missing data statistics.
2. Read the *.lmiss file.
3. Read the *.imiss file.

Minor Allele Frequency (MAF) Exercise:
Determine minor allele frequency using the entire dataset (no pop structure)
1. Produce MAF statistics.
2. Read the *.frq file.
Calculate MAF per subpopulation
1. Read the pop.phe file.
2. Produce MAF statistics using the pop.phe file.
3. Read the *.frq.strat.

Filtering Options Exercise
1. Retain markers with a genotype call rate of 85%.
   a. How many markers were removed?
   b. How many markers were retained?

Logistic Regression Exercises
1. Run a logistic regression with covariates.
2. Run a logistic regression with covariates, but hiding the covariates.
3. Run a logistic regression without covariates.
4. Run a logistic regression with covariates, hide the covariates, and correct for multiple testing.

Manhattan Plot Exercise
1. Produce a Manhattan Plot using the 'ManhattanPlot.Rmd' and the file 'withpopB.assoc.logistic'.