Investigating linkage disequilibrium in different populations

MMG1001 Assignment 2 – Group Heather (originally developed by Kaitlin Laverty)

Corresponding reading

Gabriel et.al. 2002.The Structure of Haplotype Blocks in the Human Genome. Science 296(5576):2225-2229

Additional files needed

- haplotypeblock1.txt and haplotypeblock2.txt (download from Quercus)
    - These files contain 6 SNPs each, roughly within the same haplotype block
- The group worksheet

Tools

- LDlink: a suite of web-based applications designed to easily and efficiently interrogate linkage disequilibrium in population groups
- dbSNP: a database of curated SNPs

Overview

Examining the haplotype structure of different populations can tell us a lot about the evolutionary history of our genome and can provide insights into human diseases. We will explore a few haplotype tools to see what kind of information we can derive from SNPs.

Group Worksheet

In the group worksheet, please fill out the definition for one term (table 1) and the information for one SNP (table 2). You should be prompted to log in to your UofT account to access the worksheet.

Determine if two SNPs are in linkage disequilibrium

1. Using the **LDlink tool *LDpair***, determine if the following pairs of SNPs are in linkage disequilibrium.

SNP 1	**SNP 2**	**D'**	**R^2**	**Most frequent haplotype**	**In LD?**
rs79888670	rs111679233				
rs76370881	rs61946969				
rs76370881	rs142261558				

2. Do you agree with the results from the tool?

Determine haplotypes in different populations

Use the **LDlink tool *LDhap*** to determine all possible haplotypes in various populations for the six SNPs in the haplotype block your SNP is in (from the Group Worksheet). Copy/paste the six SNP IDs into the text box (one per line) or upload the file (haplotypeblock1.csv or haplotypeblock2.csv).

3. Select the appropriate populations (click on the name of the continental population to quickly select all of the subpopulations) and write how many haplotypes are found in each of the following and the frequency of the major haplotype:
    a. All European groups:
    b. All African groups:
    c. All East Asian groups:
    d. Shared across all European, African, and East Asian groups:
    e. Are you surprised by the results? Why or why not?

Imputing SNPs

4. Using the **LDlink tool *LDpop***, check the linkage disequilibrium between the SNPs in rs1325809 and rs142261558 for the following populations:
    a. Finnish:
    b. Yoruba:
5. Using cutoff values of D'>0.95 and r^2>0.5, could you confidently impute the genotype at rs142261558 for each of the above?
6. In which additional populations could you confidently impute rs142261558 from rs1325809?

Finding tag SNPs associated with SNPs of clinical significance

Instead of making a custom SNPchip that genotypes a pathogenic SNP, we can instead impute the allele at that locus with a nearby SNP (tag SNP) with a higher allele frequency that occurs on a commercially available SNPchip.

7. Use the **LDlink tool *LDtrait*** to see if any of the following SNPs are tags for a pathogenic SNP: rs11571818, rs76370881
    a. Which one is the tag SNP for a pathogenic SNP for breast cancer?
    b. What is the ID of the pathogenic SNP?
    c. Click on the link in the ***LDpair*** column. Which allele from the tag SNP is associated with the pathogenic allele?
    d. Confirm the pathogenic allele by searching for the SNP in **dbSNP**. Why is the allele pathogenic?
    e. SNPs of clinical significance (e.g., pathogenic SNPs) often occur at a rare frequency. Why is this?

Additional questions to ponder

1. What information (from a gene we learned about in class) can we use to better define haplotype block boundaries?