Human SNP haplotypes

Statistics 246, Spring 2002

Week 15, Lecture 1

Human single nucleotide polymorphisms

The majority of human sequence variation is due to **substitutions** that have occurred **once** in the history of mankind at **individual base pairs**, SNPs (Patil et al, 2001 listed at the end, and refs therein).

It has been estimated that **> 5 million common SNPs**, each with a frequency of 10% - 50% account for the bulk of human DNA sequence difference.

Such SNPs are present in the human genome about **1 in every 600** base pairs.

Alleles making up **blocks** of such SNPs in close physical proximity are often correlated, and define a limited number of **SNP haplotypes**, each of which reflects descent from a single, ancient ancestral chromosome.

The Daly *et al* (2001) data set

This consists of 103 **common SNPs** (>5% minor allele frequency) in a 500 kb region implicated in Crohn disease, genotyped in 129 trios (mom, pop, kid) from a European derived population, giving 258 transmitted and 258 untransmitted chromosomes.

Studies to date have revealed **great variability in local haplotype structure**: the relative contributions of mutation, recombination, selection, population history, and stochastic events seems to vary unpredictably. Some haplotypes extend only a few kb, while others extend for > 100 kb.

Here is some evidence from Figure 1 of Daly *et al*, 2001. **Linkage disequilibrium** (LD) between an arbitrary marker (#26 in a, #61 in c, see *) and every other marker in the data set is indicated, using the **normalized** association measure D'= (ad-bc)/(a+c)(c+d) of LD.

Note the noisiness of the plot.