

**Haplotyping Exercises**

1. To study the role of the LAZY gene in susceptibility to tardiness, 5 normal individuals and 5 controls were genotyped for 3 markers within the gene.

The following genotypes were observed:

Genotypes			Number of Observations	
Marker 1	Marker 2	Marker 3	Cases	Controls
1/1	1/1	1/1	2	0
1/2	1/2	1/2	3	3
1/1	2/2	2/2	0	2

Using an E-M algorithm, haplotype frequencies for the combined data set were estimated as  $p_{111} = 0.35$ ,  $p_{122} = 0.35$ ,  $p_{222} = 0.15$  and  $p_{211} = 0.15$ .

- a) Describe and apply a likelihood ratio test to compare haplotype frequencies between cases and controls.
  - b) Describe how the “haplotype trend” regression method of Zaykin et al. could be applied to this data. Include the expected dosage of the 4 haplotypes for individuals in each of the three observed genotype classes.
2. Consider the following set of observed genotypes for 4 nearby markers:

Marker1	Marker2	Marker3	Marker4	Count
1/1	1/1	1/2	1/2	10
1/2	1/2	1/2	1/2	10
1/2	1/2	2/2	2/2	10

Use an E-M algorithm to estimate haplotype frequencies for these 4 markers. To facilitate calculation, use the following procedure:

- a) Estimate haplotype frequencies for markers 1 and 2 only; list observed haplotypes.
- b) Estimate haplotype frequencies for markers 3 and 4 only; list observed haplotypes.
- c) Finally, estimate haplotype frequencies for the entire region. To do this, first combine haplotype lists from steps a) and b) to create a list L of likely haplotypes for the entire region. When carrying out this final step assume that all haplotypes not included in list L have frequency zero, so that you only have to update frequencies for haplotypes listed in L using the E-M algorithm.