

Algorithms in Genome Research
Winter 2023/2024

Exercises

Number 10, Discussion: 2024 January 26

- Investigate: What is the difference between linkage and association?
- Let the following haplotype matrix be given (circles represent the presence of a mutant allele):

	1	2	3	4	5	6	7	8	9	10
A	—O—	—————		O—O—O—					O—O—	
B			O—				O—			
C	—O—O—				O—O—			O—O—O—		
D			O—					O—		
E	—O—O—					O—			O—	
F				O—			O—		O—	

- Find the maximal regions around each segregating site, not violating the four-gametes test.
 - Draw the local trees for these regions.
 - Assume that individuals *C* and *E* are the cases, the other the controls. Which of the segregating sites show highest evidence for association with the disease?
- Discuss: Can pedigree (family) information help in the analysis of whole-genome association studies?
 - Give general formulas for the following questions. If this is difficult, enumerate the solutions for small examples.
 - For n biallelic sites, represented by the columns of a $k \times n$ binary haplotype matrix, how many *different* haplotype sequences (rows) are at most possible?
 - If the haplotypes come in blocks of 10 sites each, 2 variants each, how does this decrease the number of *different* haplotype vectors?
 - For l founder sequences and m recombination hot spots, how many haplotype vectors are possible (under the assumption that recombinations occur only at hot spots)?
 - For k haplotype sequences, what is the maximum number $n_{\max}(k)$ of *different* configurations at segregating sites (columns) such that the four-gametes test does not fail?