

## Assignment 1a

Assignment 1 in computer class sessions (and also at own time) is learning to use PHASE (or fastPHASE) for genotype phase inference. This is an important practical skill, as current high-throughput genotyping technologies (for example SNPs from diploid individuals) determine which two alleles (for example nucleotides at a SNP site) are present, but not the haplotype information: which *combinations* of alleles (nucleotides) are present on each of the two chromosomes. Knowledge of the haplotypes carried by individuals is important for many purposes, including linkage disequilibrium mapping and inference of population evolutionary history because inheritance, in broad sense, is a relevant concept as transmitting large chromosomal segments.

Assignment 1 consists of parts 1a and 1b. Before the first computer class session you are supposed to perform part 1a and submit the result to the Student page (click the page link and you'll find more information). Part 1b starts during the first session and more information will be given.

**Now you are supposed to find out what is the relevant algorithm(s) in PHASE, i.e. how does the program perform the inference.**

**Write a short (~ one page) report. You are not supposed to write long referee about PHASE, just show that you have tried to understand the algorithm(s). You don't have write formulae in you report (you can do so, if you like).**

The first relevant paper (link in the course page: **PH1.pdf**) is

Stephens, M., Smith, N., and Donnelly, P. (2001). [A new statistical method for haplotype reconstruction from population data](#). *American Journal of Human Genetics*, 68, 978-989.

From the paper (**PH2.pdf**) you may get additional information:

Stephens, M., and Donnelly, P. (2003). [A comparison of Bayesian methods for haplotype reconstruction from population genotype data](#). *American Journal of Human Genetics*, 73:1162-1169.

Have a look also in the most recent paper (**PH3.pdf**) in which the a new version of PHASE is presented :

Scheet, P and Stephens, M (2006). [A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase](#). *Am J Hum Genet* 78, 629-644.