

Homework 1

Topics in Bioinformatics

Important dates:

- Submit report and presentation files before 9 November 2016, 24:00h

Marks:

- Programming part – 30 points

Evaluation:

- For report, your work will be evaluated based on the accuracy and completeness of answers and report structure.

Instruction:

- Form a group of 2-3 persons and complete the homework.
- A report needs to be submitted in electronic format via the website by the deadline. Please note that the submission system will be closed automatically.
- Compress all files into ONE zip file and submit to:

http://www.student.montefiore.ulg.ac.be/~GBIO009-1/GBIO0009_sys/index.php

Homework on GWAS

Programming

Use R and/or PLINK to analyze the provided data and answer the following questions using a report format (introduction, discussion and conclusion sections). Your source codes need to be submitted together with your report file.

- Dataset
 - Download the HAPMAP3 data
 - Extract these 2 populations: CEU and YRI
 - For case – control study, assume CEU population as case and YRI population as control
- Knowing the data – explain what are in the data files, for example:
 - How many individuals? Cases? Controls?
 - How many SNPs?
 - How many chromosomes? XY chromosomes are included?
- Quality control – apply the quality control processes to the data and explain the purpose of each process, for example:
 - HWE test
 - Missing genotype filtering
 - LD pruning
 - How many individuals and SNPs remain after each QC process?
- Analysis results – do association test and explain the results, for example:
 - A list of top significant SNPs
 - Manhattan plot
 - QQ plot
- Comparison – you may compare based on your own design, for example:
 - The association test with QC vs the association test without QC
 - Association test for different SNP encoding schemes