

# Project: population and family-based analysis

The project uses data from the FAMuSS study (Functional SNPs Associated with Muscle Size and Strength). The population data are available as tab-delimited text file. Background information can be found in Thompson et al. 2004. In addition, simulated family-based data is available as pedigree and map file (PLINK format). It contains data on 300 families with parents selected from the Caucasians in the population based study.

## Part A (mandatory)

Possibility to improve on HW3 and HW4, in particular paying attention to correct interpretation of results (check slides with feedback on HWs 3&4 for further guidance)

If going for Part B as well, make the following changes to Part A

HW3 Q6: Use trait NDRM.CH instead of pre.BMI>25

HW4 Q4: Set control parameter usesurrogate=0

## Part B (optional)

*Main question:* What (genetic) factor(s) play a role in the trait NDRM.CH?

*Additional guiding questions*

Q1: Concretely adopt QC measures based on the data quality checks done previously in the population data, i.e. decide which (genetic) variables / individuals to remove

Q2: Is additional QC needed and/or possible when using families

Q3: Discuss the association results obtained previously in the population data. Discuss consequences of (not) correcting for multiple testing and its relevance. Is it important to adjust for confounding factors?

Q4: Use GenABEL to generate association results in the family data. Discuss consequences of (not) correcting for multiple testing and its relevance. Is it important to adjust for confounding factors?

Q5: How do the two previous analyses compare? Why are there differences? What are the advantages/disadvantages of using families?

Q6: What could the trees built previously in the population learn you about genetic effect modifiers?

Bonus question: Perform an interaction study on either version of the data using the packages seen in class