

Practicals Bioinformatics 2011-2012

Olivier Stern olivier.stern@ulg.ac.be
Tom Cattaert tom.cattaert@ulg.ac.be

20 December 2011: Feedback homeworks 3&4

General

- These slides offer class feedback on which problems occurred, besides individual feedback in corrected copies
- Should answer the questions rather than just produce output
- Possibility to improve on homeworks 3 and 4 in Part A of project assignment (mandatory!)
- Pay particular attention to interpretation of results
- As interpretation was not yet taken into account, grades may both increase and decrease

Homework 3 question 1

- Read in data correctly instead of discarding difficult line from input
- Check the meaning of Hardy-Weinberg equilibrium
- Double check code which did not work for some students

Homework 3 question 2

- Include all valid SNP variables
- Clarification: let us agree on the following list

```
fms[,c(2:29,31:52,54:110,115:164,166:226)]
```

- The same list is used in question 6 and in Part B of the project

Homework 3 question 3

- Do not confuse genotyping percentage with association results
- Do not adjust for other SNPs
- Clarification: when adjusting adjust for all three predictors at once
- Best use packages seen in class

Homework 3 question 4

- Take into account the correct genetic model asked for
- If testing is asked, p-values should be reported
- Clarification: analysis stratified by gender and analysis within gender strata are exactly the same thing, hence only one analysis is needed besides the original analysis in the complete sample

Homework 3 question 5

- OR is always comparing one category to another category
- Restricting to homozygotes is not necessary, as the answer can be seen from standard output

Homework 3 question 6

- Certainly here it is not sensible to restrict to homozygotes as all genetic models are asked
- No sensible output from WGassociation, so try something else e.g. one by one association or GenABEL package

Homework 3 question 1

- Code did not work for some students so please double check
- Unrelated individuals should be in different families
- Use all individuals but include pre.BMI>25 as a trait
- Not needed to convert letters into 1 2 3 4
- Could also generate info file using dummy positions

Homework 3 question 2

- Best take color scheme and values consistently as D' or r^2
- You might discuss advantages and disadvantages of both
- Do not forget to perform haplotype association analysis with trait pre.BMI>25

Homework 3 question 3

- Pruning was asked and is relevant also for small trees
- rpartOrdinal is package for ordinal responses rather than ordinal predictors, how to achieve ordinal predictors was discussed in class
- Show results of categorical/ordinal version consistently
- Be careful about discussion of link between genotypes and ordinal numerical values

Homework 3 question 4

- Do not interpret `yval` as percentage variation explained
- Include both `actn3` and `resistin` genes
- Sensible output only when setting control parameter `usesurrogate=0`

Homework 3 question 5

- Remind that analysis is not reproducible without random seed (in fact also for pruning in previous questions)
- Results should be shown corresponding to random seed in code
- Make sure to at least put all relevant output answering the questions in your report, but also not more than that
- Clarification: asked for a single analysis restricted to Caucasians
- Report genotype missingness rates before and after `na.roughfix()`

Homework 3 question 6

- Not just repetition of question 5 on whole data nor on strata
- Think carefully about correct way to deal with problem stated