|  |  |
| --- | --- |
| GBIO0002 | 1 |
|  |  |

**Homework 1**

Genetics and bioinformatics

**Important dates:**

* Submit report /presentation files before 6 November 2018,
* The presentation will be on 20 November 2018

**Marks:**

* Part 1- 15 marks (select either Q&A or literature-based homework)
* Part 2- 15 marks (select either Q&A or literature-based homework)

**Evaluation:**

* For the Q&A homework your work will be evaluated based on the accuracy and completeness of answers. Write your answers in the form of a report.
* For the literature-style homework, evaluated based on the completeness of your slides, your presentation skills, your understanding of the matter, and your answers to questions in class (see the information given during the introductory class). No report is needed. Recall the document “critical evaluation of a report/paper”.

**Instruction:**

* Form a group of 2-3 persons and complete the homework in both parts (1&2). Make sure there are about 8 groups maximum.
* For the presentation, you have to present your selected paper within 15 minutes. Everybody in your group should present something.
* A report / slides presentation(whatever is applicable – see above) 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 before submission.

|  |  |
| --- | --- |
| GBIO0002 | 2 |
|  |  |

**Part 1:** Genetics and DNA sequencing.

**Q&A homework**

Select only ONE task from the list (A, B, or C).

**A**: Answer the following questions using a report format (introduction, discussion andconclusion sections).

* List the key features of the DNA structure.
* Explain which DNA feature(s) is (are) important for its genetic information encoding ability (think in terms of encoding the information but also decoding).Explain which feature(s) is (are) important for protecting the heredity information taking into account the context of the genome, cell structure and gene expression steps (DNA→mRNA →protein).

**B**: Answer the following questions using a report format (introduction, discussion andconclusion sections).

* Explain the Sanger DNA sequencing method. What are the main drawbacks of this method?
* Explain the cloning approach when sequencing large DNA fragments. What are the pro and cons of the method?
* Explain in details the bridge PCR technology?

**C:**Answer the following questions using a report format:

* What are genetic molecular markers?
* Why are they useful for?
* What are their utilities in human medicine (illustrate using 2-3 examples)?

|  |  |
| --- | --- |
| GBIO0002 | 3 |
|  |  |

**Literature-style homework**

Select only ONE of the following papers for presentation (cfr. info given at the first class). Your presentation should cover the objective, method, results (if available), literature references (if needed), your own discussion, and your own conclusions.

* Van Dijk, Erwin L., et al. "Ten years of next-generation sequencing technology." Trends in genetics 30.9 (2014): 418-426.
* Parsons, B.L., 2018. Multiclonal tumor origin: evidence and implications. *Mutation Research/Reviews in Mutation Research*.
* Zhong, L., Liu, Y., Wang, K., He, Z., Gong, Z., Zhao, Z., Yang, Y., Gao, X., Li, F., Wu, H. and Zhang, S., 2018. Biomarkers: paving stones on the road towards the personalized precision medicine for oral squamous cell carcinoma. *BMC cancer*, *18*(1), p.911.

|  |  |
| --- | --- |
| GBIO0002 | 4 |
|  |  |

**Part 2:** Genome-wide association studies - analytics.

**Q&A homework**

**Background information about the experiment**

Data at our disposal consisted of a small region of about 2.5 Mb of chr1. In total 833 SNPs were typed on 2500 people. The missingness rate was 95%. Sex, age, two quantitative traits (qt1 and qt2) and one binary (bt) trait were available for analysis. We performed corresponding associationanalyses among the SNPs and trait qt1 and generated several plots during the analyses (which mimic analyses shown in the practicals). Essential code is given as an appendix.

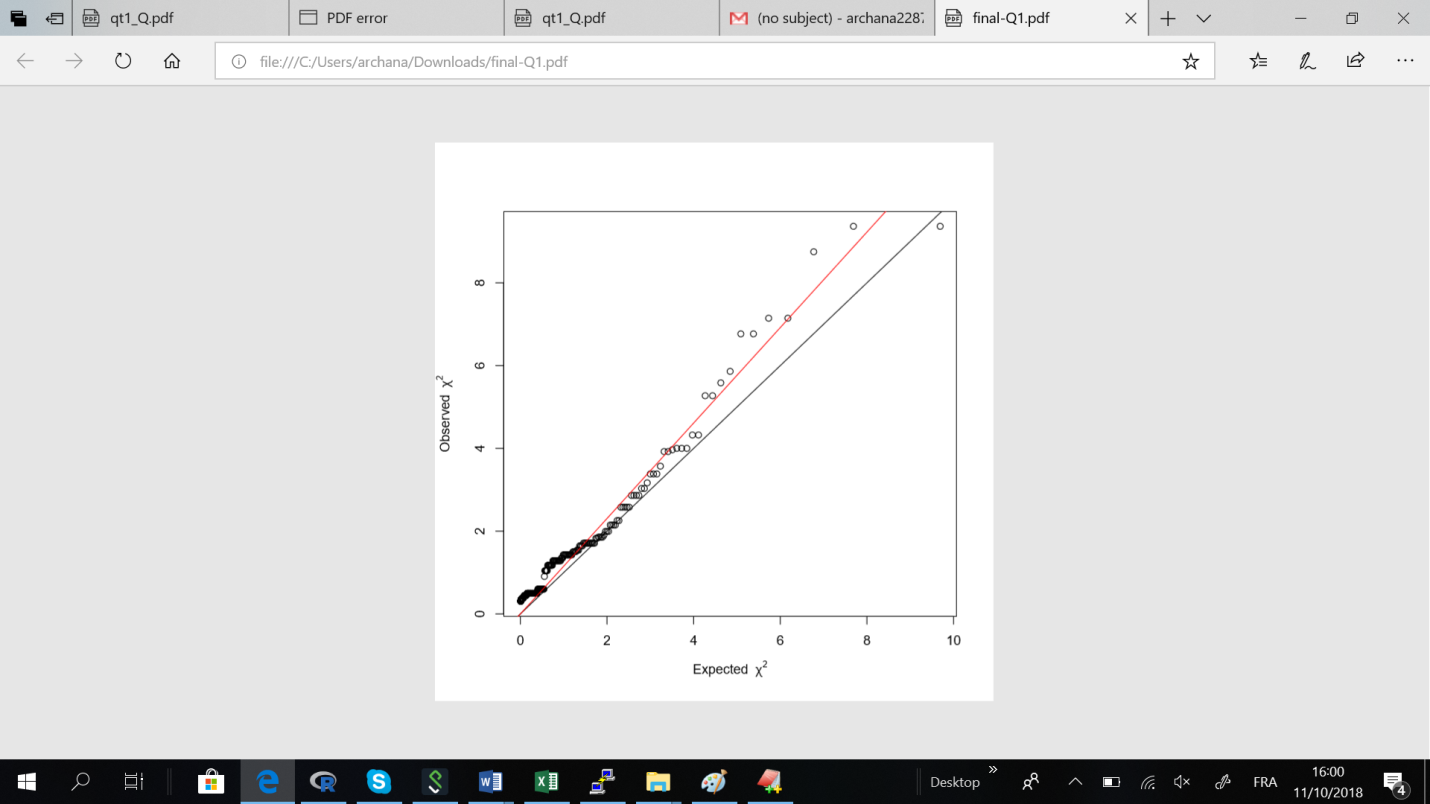
**Draft a small report around the following questions**:

1) What are main characteristics of the data?

2) What are the main characteristics of a genome-wide association study (components of the analysis)?

3) To which component of a genetic association study do the following plots (Figure 1) relate?

4) What do these plots (Figure 1) present? What is their interpretation? Do the estimated lambda’s raise concern? Do the results require actions? If so, which ones?



Estimated lambda = 0.95(control) Estimated lambda = 1.15(cases)

**Figure 1**

5) To which component of a genetic association study does Figure 2 relate?

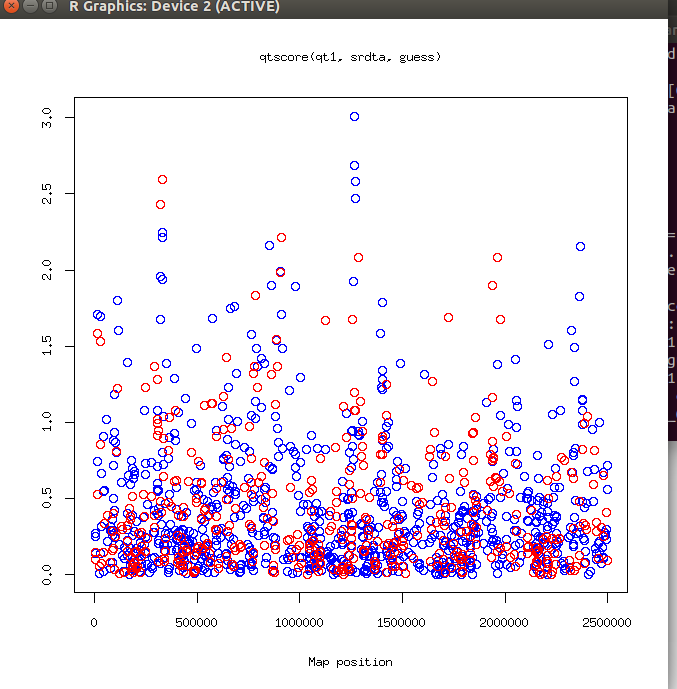
6) What is the test statistic used in the association analysis for q1? Please describe in more detail.

7) What does Figure 2 represent? What is its interpretation (meaning of coloring, x-axis, y-axis, )?

8) Were confounding factors taken into account? If so, which ones? What is a confounding factor to an association of interest?

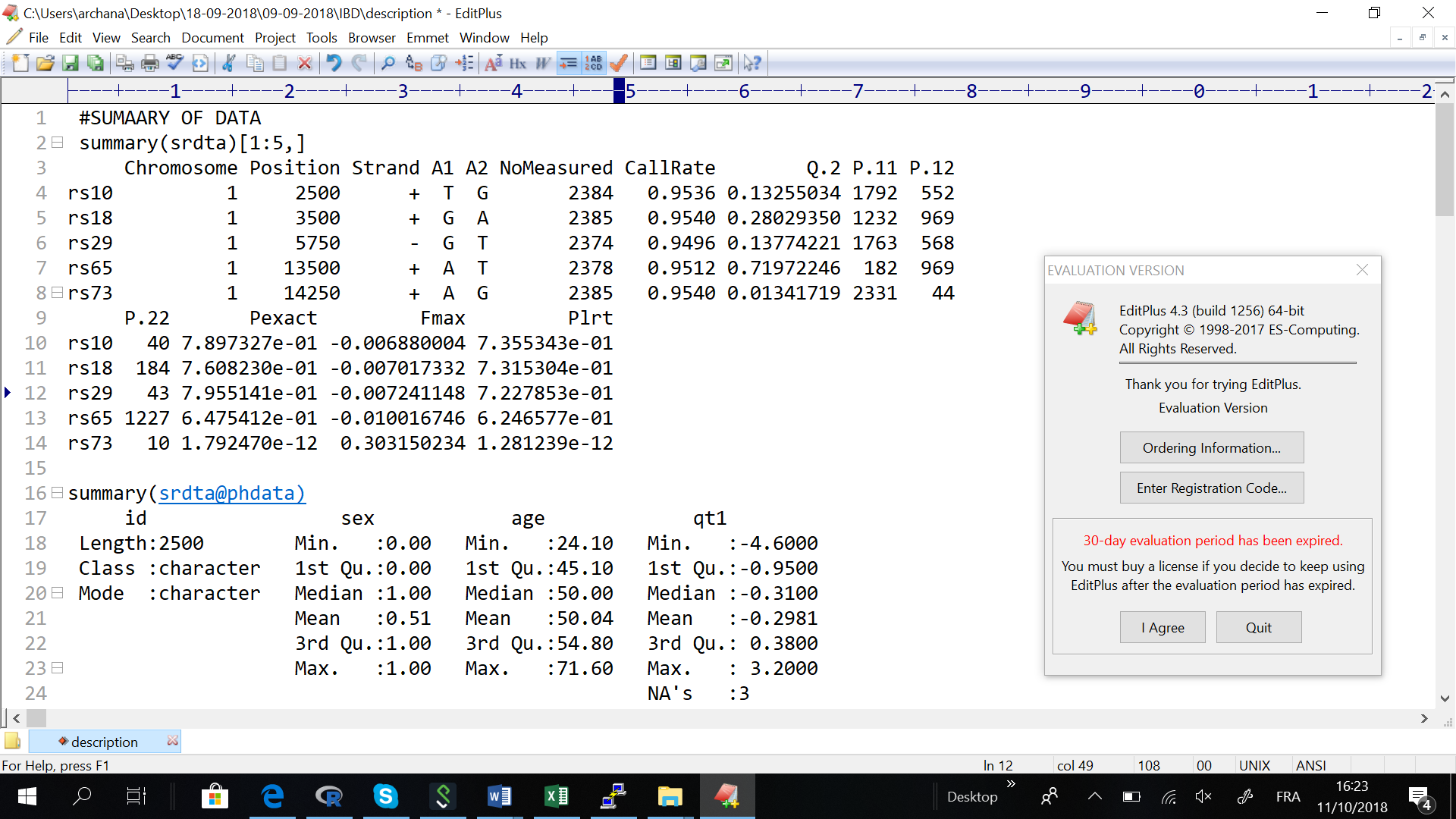
9) What is the effect of multiple testing? Was it performed here?

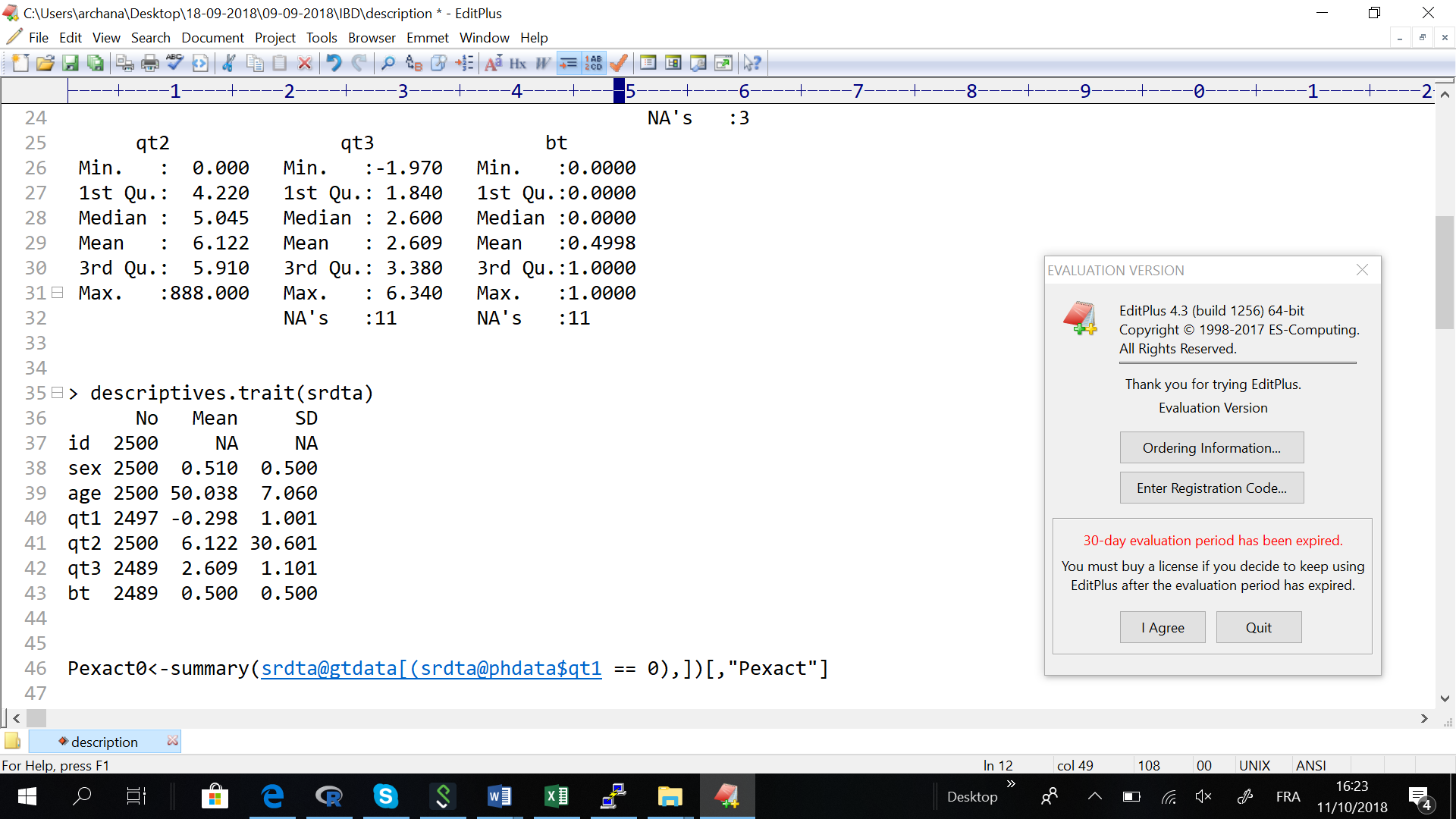
10) What would be the next steps towards interpretation?

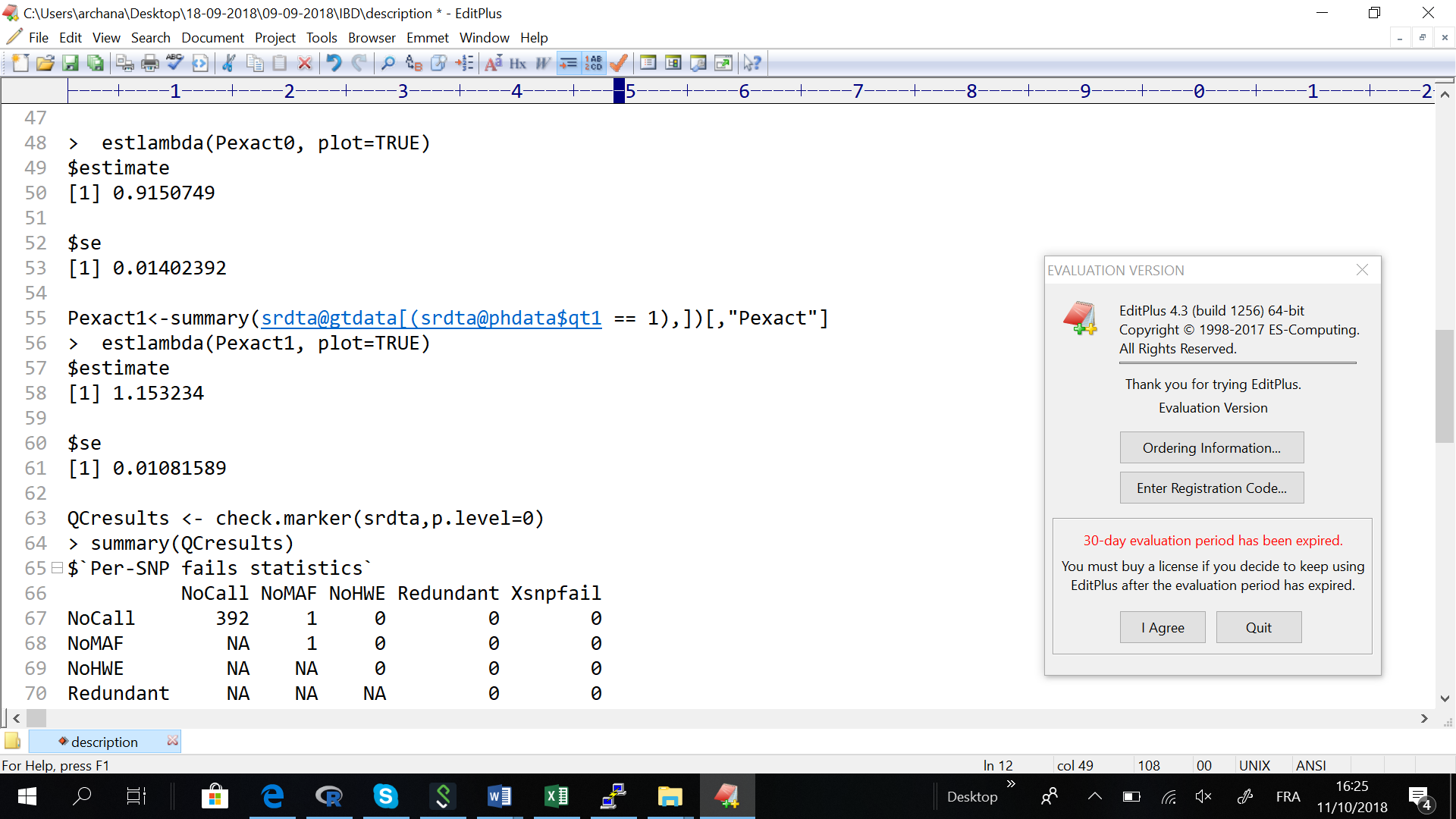


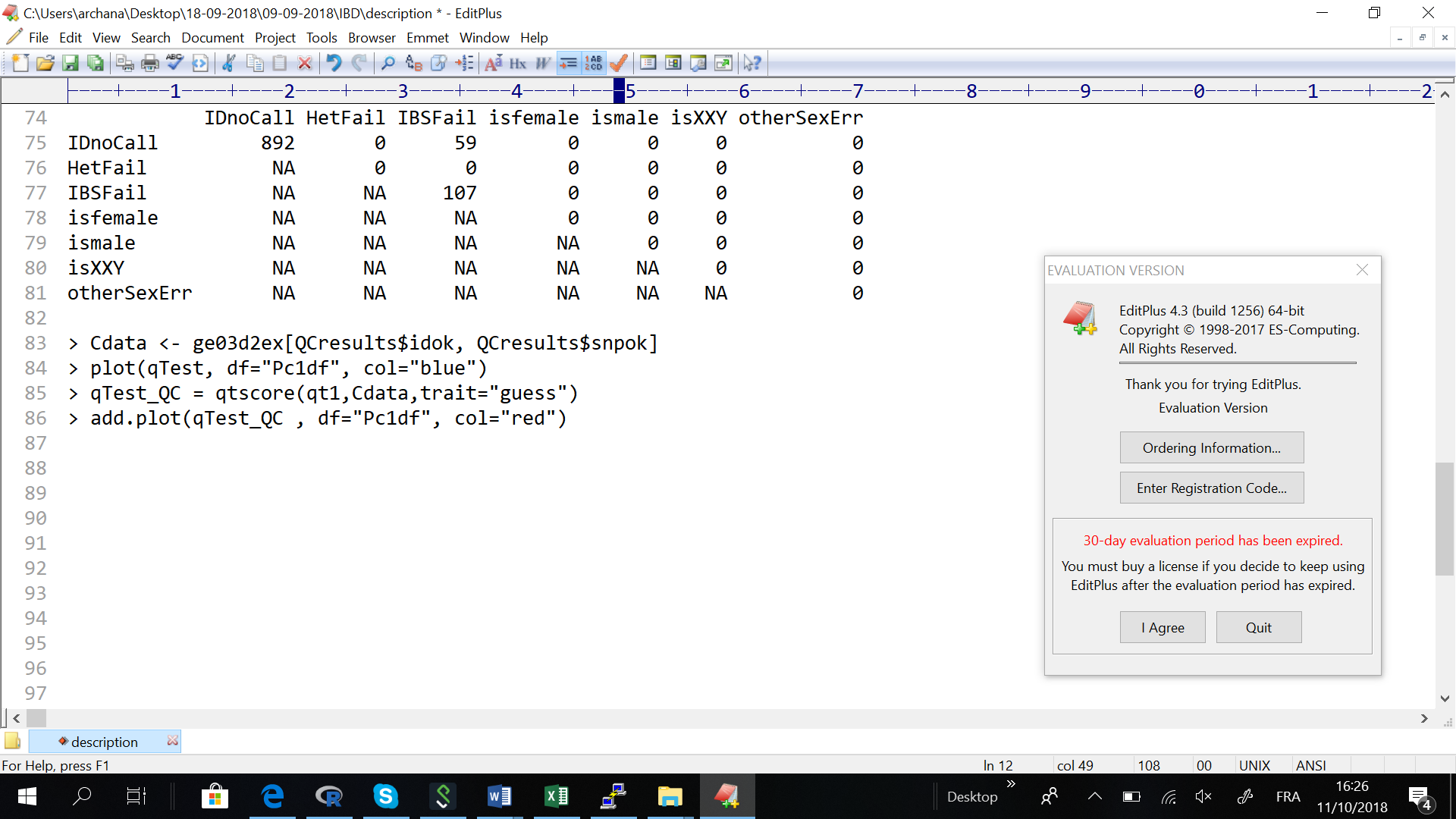
**Figure 2**

**Appendix**









**Literature-style homework**

Select only ONE of the following papers for presentation (cfr. info given at the first class). Your presentation should cover the objective, method, results (if available), literature references (if needed), your own discussion, and your own conclusions.

1. Savage 2018 - GWAS intelligence   
  
2. Michailidou 2017 - GWAS breast cancer   
  
3. van der Harst 2018 - GWAS coronary artery disease   
  
4. Saint Pierre 2014 - rare variants and common diseases   
  
5. Colonna 2009 - population structure