



You are logged in as **JOAN FRANCESC GILABERT NAVARRO** ([Log out](#))

cat esp eng **other** ▼

[La Factoria](#)

[Aula Global Help](#)



My home ► 2015-31032-T1 ► General ► Project description (2nd part)

### Navigation



My home

- Site home
- Site pages
- My profile
- Current course

### Administration



Course  
administration

My profile  
settings

## 2015-31032-T1 IEO. Infomation Extraction from OMICS Technologies

# Project description (2nd part)

The final result of the project should be a report in the form of a scientific publication re-analyzing the data set that you have chosen, with a supplementary material in form of a web page, containing all the scripts, data and figures that support the contents of the report and that make the entire analysis **reproducible**.

## Technical requirements

The report should be written in LaTeX (see <http://www.latex-project.org>) and therefore one milestone should be that you get acquainted with the basics of elaborating a document using this text processing system.

If you work in a Linux system, LaTeX should be already installed, otherwise try at all times to install it through the package installation system of your Linux distribution (because of the many files involved). If you work in a Windows or Mac OS system, then I recommend you to install the TeXLive distribution (<http://www.tug.org/texlive>).

Once you have got LaTeX installed in your computer you can try to process and modify the template document you will find under the folder 'latexTemplate' link from the moodle site of the course. That should allow you to quickly grasp the very basics steps of creating a LaTeX document, but you can find lots of introductory material to LaTeX by Googling "latex beginner".

For actually formating your report, you should follow the template guidelines of GSA Genetics at <http://www.genetics.org/content/prep-manuscript>. In particular, you should use the template provided, downloading it by first going to <http://bit.ly/1zXBpRO>, then follow "OPEN AS TEMPLATE", then follow "PROJECT" and finally click on "DOWNLOAD AS ZIP"). It is important that you follow the proposed organization of the manuscript into Title, Authors and Affiliations, Abstract, Introduction, Materials and Methods, Results and Discussion, Acknowledgments and References. Please embed figures and tables in the appropriate places along the manuscript. For further questions about the typesetting of the article please read

carefully the GSA Genetics instructions at <http://www.genetics.org/content/early/2016/05/20/1015016> and prepare a manuscript. Consider having most of supporting figures and tables in the Rmarkdown document of the Supplementary Material and only including in the main article those that are most relevant to the final results.

In your report you should **not** target a full-length standard research article. Nor should you copy & paste anything from any other article. Your report should include at least one figure and one table, but just one figure and one table in the main article is **already OK** and can be as short as 3 or 4 pages and please no longer than 7 or 8.

The materials you have to submit should be a tar ball (or zip file) containing a single PDF file of the report including the main text and all figures and tables and an HTML web page, and the source R markdown (.Rmd) file describing the supplementary material along with the data. The supplementary material should allow one to reproduce the results described in the main article. Ideally, you should build up on the supplementary material that you started developing for the first part of the project.

## Content requirements

As for the particular contents that you should develop in this second part of the project, they should include:

1. Everything from the first part of the project, or the corresponding one if you changed the data set, i.e., description of the biological context of the data, QA and batch effect identification and removal (if necessary) and normalization.
2. Differential expression analysis. For a particular contrast of interest, show how many genes change their expression, including a histogram of the p-value distribution and a volcano plot.
3. Functional analysis, either by classical functional enrichment with the Fisher's exact test, or some other approach that you think it provides more sensible results.
4. Interpretation of the results. Draw some conclusion from the functional analysis in the context of the data and of the biological question that you set out to answer in the introduction. Your results **do not** have to necessarily reproduce the ones contained in any article analyzing the original data, although if that is possible, it will be great for you as your interpretation will be greatly simplified.

## Deadline

Using the submission link you will find at the top of the IEO moodle site, submit the tarball (or zip file) of the directory structure containing the source R Markdown file, data, figures, resulting HTML and PDF of the report by Sunday 19th of June, each fraction of 24h delay will be penalized with 1 point (out of the 10 possible). There is a maximum file size submission limit of 400Mb. If your zip file is larger, remove the cache directory created by knitr (typically called 'cache' or '.cache') and the raw data from the file.

## Rubric for the project report

The rubric below contains aspects that will be taken into account when grading the report of the project. Since the report you have to submit takes the form of a

scientific article) these criteria should help you to work on items that will be important whenever you write a scientific article in the future. Please assess these aspects on the report you write before submission and fix those that you think that do not meet the criteria.

1. Does the paper have the sectioning from the GSA Genetics author guidelines? (Title, Authors and Affiliations, Abstract, Introduction, Materials and Methods, Results, Discussion, Acknowledgments and References)
2. Is the text written in a cohesive way? Does the story about the data flows naturally through the article?
3. Are processing steps and analysis decisions well justified?
4. Are figures labeled and referred to by number in the text?
5. Is the total number of samples and samples per group, reported at each stage of the analysis?
6. Is the total number of genes reported at each stage of the analysis?
7. Are figure and table captions descriptive enough to stand alone?
8. Are axes in plots labeled in plain language and large enough to read?

Last modified: diumenge, 22 maig 2016, 10:21

You are logged in as **JOAN FRANCESC GILABERT NAVARRO** (**Log out**)

Legal advice