

Tutorial 2 - Differential Analysis and Design

Lieven Clement

The result of a quantitative analysis is a list of peptide and/or protein abundances for every protein in different samples, or abundance ratios between the samples. In this chapter we will extend our generic workflow for differential analysis of quantitative datasets with more complex experimental designs.

We will start with assessing the impact of sample size. Then we will introduce the concept of blocking and we will continue with the analysis of a subset of the mouse diet data.

Breast cancer example

Eighteen Estrogen Receptor Positive Breast cancer tissues from patients treated with tamoxifen upon recurrence have been assessed in a proteomics study. Nine patients had a good outcome (or) and the other nine had a poor outcome (pd). The proteomes have been assessed using an LTQ-Orbitrap in data dependent acquisition (DDA) mode and the Thermo output .RAW files were searched with MaxQuant (version 1.4.1.2) against the human proteome database (FASTA version 2012-09, human canonical proteome).

The raw data can be found in the folder dda/cancer after downloading and unzipping all data locally. [Download data](#)

Preprocessed data are also already stored as QFeatures objects. They can be loaded in R using readRDS or directly with the launchMsqrob2ModelingApp(). So you only have to read the data and you can start with section 5 Data exploration in the basic script or in the GUI.

```
qf <- readRDS(url("https://github.com/statOmics/PDA-DIA/raw/refs/heads/main/data/cancer3x3.rds"))
```

Three QFeatures objects are available (you can simply modify the url above replacing the filename):

1. For a 3 vs 3 comparison: cancer3x3.rds
2. For a 6 vs 6 comparison: cancer6x6.rds
3. For a 9 vs 9 comparison: cancer9x9.rds

So you can skip the preprocessing to save some time. This can be done in a GUI that starts from a QFeatures object stored in `rds` format: `launchMsqrob2ModelingApp()`. Alternatively you can do the full data processing by adjusting the script [cptac Rmarkdown script](#) or performing the analysis using the [msqrob2GUI](#).

Blocking: Mouse T-cell example

(Duguet et al. 2017) compared the proteomes of mouse regulatory T cells (Treg) and conventional T cells (Tconv) in order to discover differentially regulated proteins between these two cell populations. For each biological repeat the proteomes were extracted for both Treg and Tconv cell pools, which were purified by flow cytometry. The data in `dda/mouseTcell` are a subset of the data [PXD004436](#) on PRIDE.

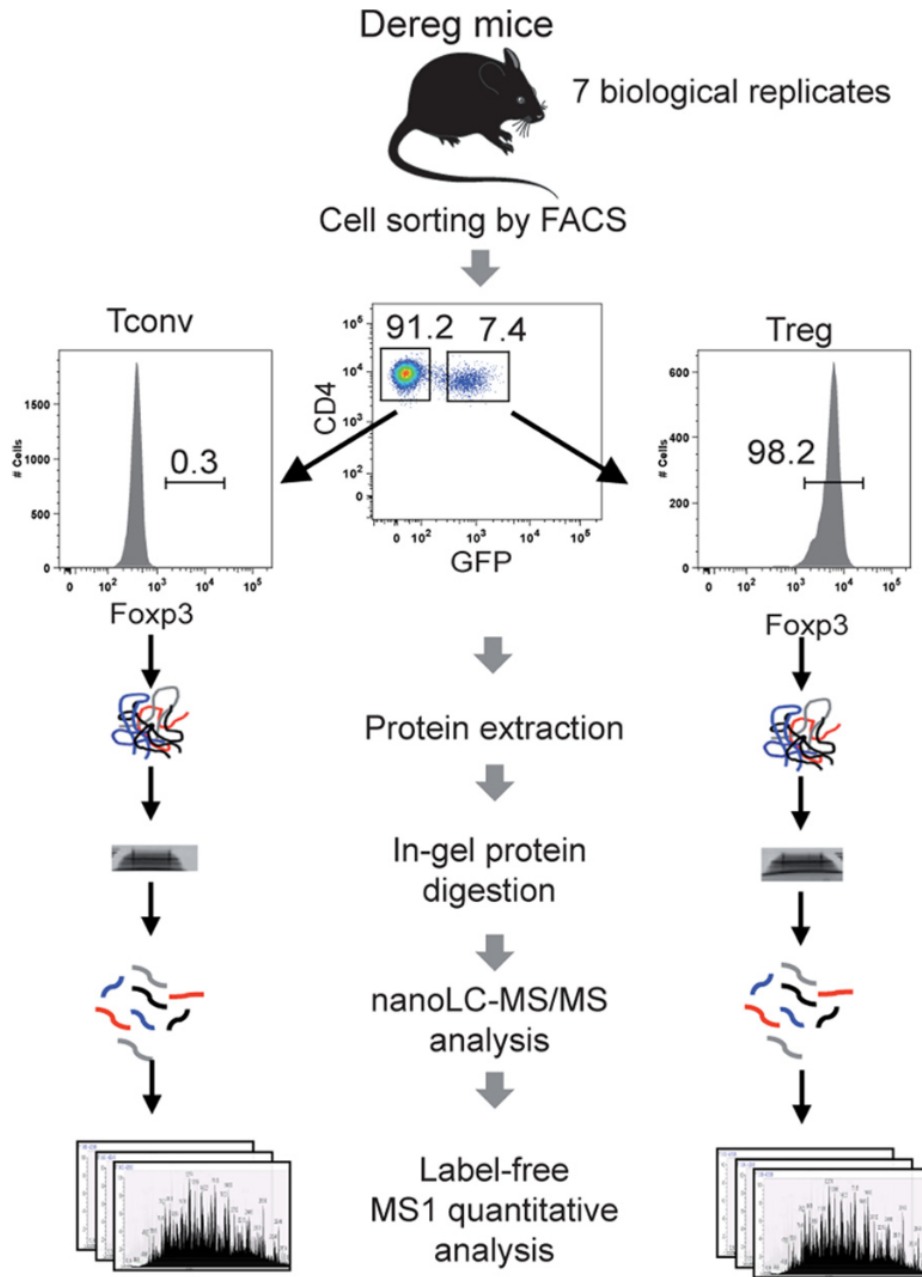


FIG. 1. Label-free quantitative analysis of conventional and regulatory T cell proteomes. General analytical workflow based on cell sorting by flow cytometry using the DEREG mouse model and parallel proteomic analysis of Tconv and Treg cell populations by nanoLC-MS/MS and label-free relative quantification.

Figure 1: Figure 2. Design Mouse Study

The proteomes have been assessed using data dependent acquisition (DDA) mode.

Three subsets of the data are available:

- peptidesCRD.txt: contains data of Tconv cells for 4 bio-repeats and Treg cells for 4 bio-repeats
- peptidesRCB.txt: contains data for 4 bio-repeats only, but for each bio-repeat the Treg and Tconv proteome is profiled.
- peptides.txt: contains data of Treg and Tconv cells for 7 bio-repeats

Users using R-scripts can import the data using following links:

```
https://github.com/statOmics/msqrob2data/raw/refs/heads/main/dda/mouseTcell/peptidesCRD.txt  
https://github.com/statOmics/msqrob2data/raw/refs/heads/main/dda/mouseTcell/peptidesRCB.txt  
https://github.com/statOmics/msqrob2data/raw/refs/heads/main/dda/mouseTcell/peptides.txt
```

Users with the GUI can find the datasets in directory dda/mouseTcell after downloading and unzipping the data ([Download data](#)).

1. How would you analyse the CRD data?
2. How would you analyse the RCB data?
3. Try to explain the difference in the number of proteins that can be discovered with both designs?

Block design with multiple factors: Heart example

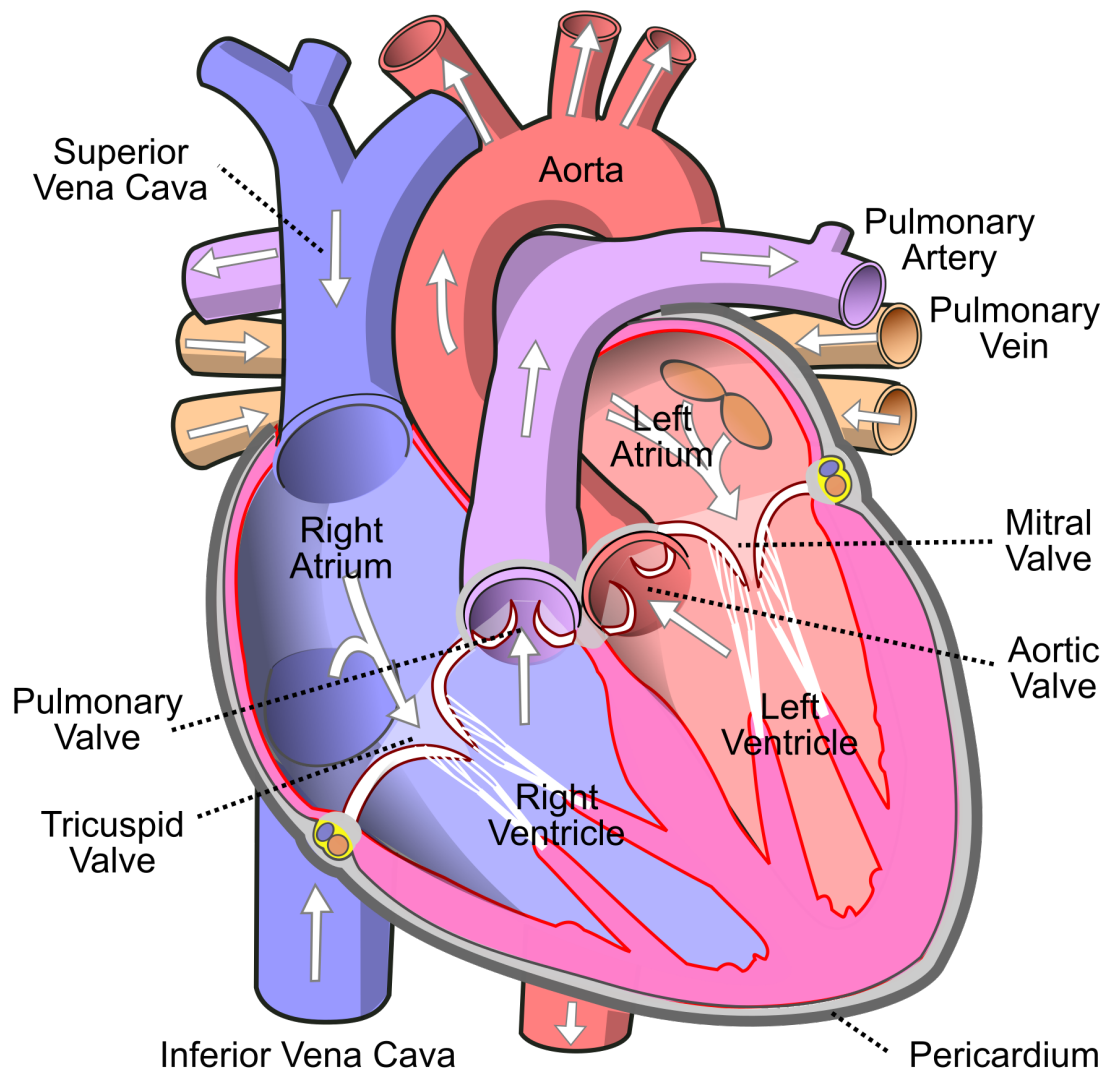


Figure 2: Figure 3. Heart

Researchers have assessed the proteome in different regions of the heart for 3 patients (identifiers 3, 4, and 8). For each patient they sampled the left atrium (LA), right atrium (RA), left ventricle (LV) and the right ventricle (RV). The data are a small subset of the public dataset [PXD006675](#) on PRIDE.

Users using R-scripts can import the data using following links:

`https://github.com/statOmics/msqrob2data/raw/refs/heads/main/dda/heart/peptides.txt`

Users with the GUI can find the datasets in directory dda/heart after downloading and unzipping the data ([Download data](#)).

Suppose that researchers are mainly interested in comparing the ventricular to the atrial proteome. Particularly, they would like to compare the left atrium to the left ventricle, the right atrium to the right ventricle, the average ventricular vs atrial proteome and if ventricular vs atrial proteome shifts differ between left and right heart regions.

1. Adjust the script [cptac Rmarkdown script](#) or perform the analysis using the [msqrob2GUI](#).
2. Redo the analysis but use median normalisation: `diff.median`, what do you observe? Explain

Duguet, Fanny, Marie Locard-Paulet, Marlène Marcellin, et al. 2017. “Proteomic Analysis of Regulatory t Cells Reveals the Importance of Themis1 in the Control of Their Suppressive Function.” *Molecular & Cellular Proteomics* 16 (8): 1416–32. <https://doi.org/10.1074/mcp.M116.062745>.