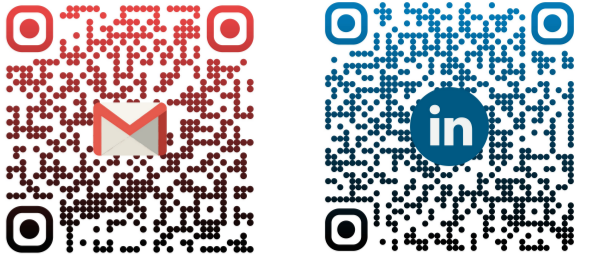


Statistical methods for the analysis of longitudinal data in quantitative proteomics



Hélène BORGES, T. BURGER, V. BRUN
EDyP team, IRIG/DS/BGE laboratory – UGA



1 Introduction

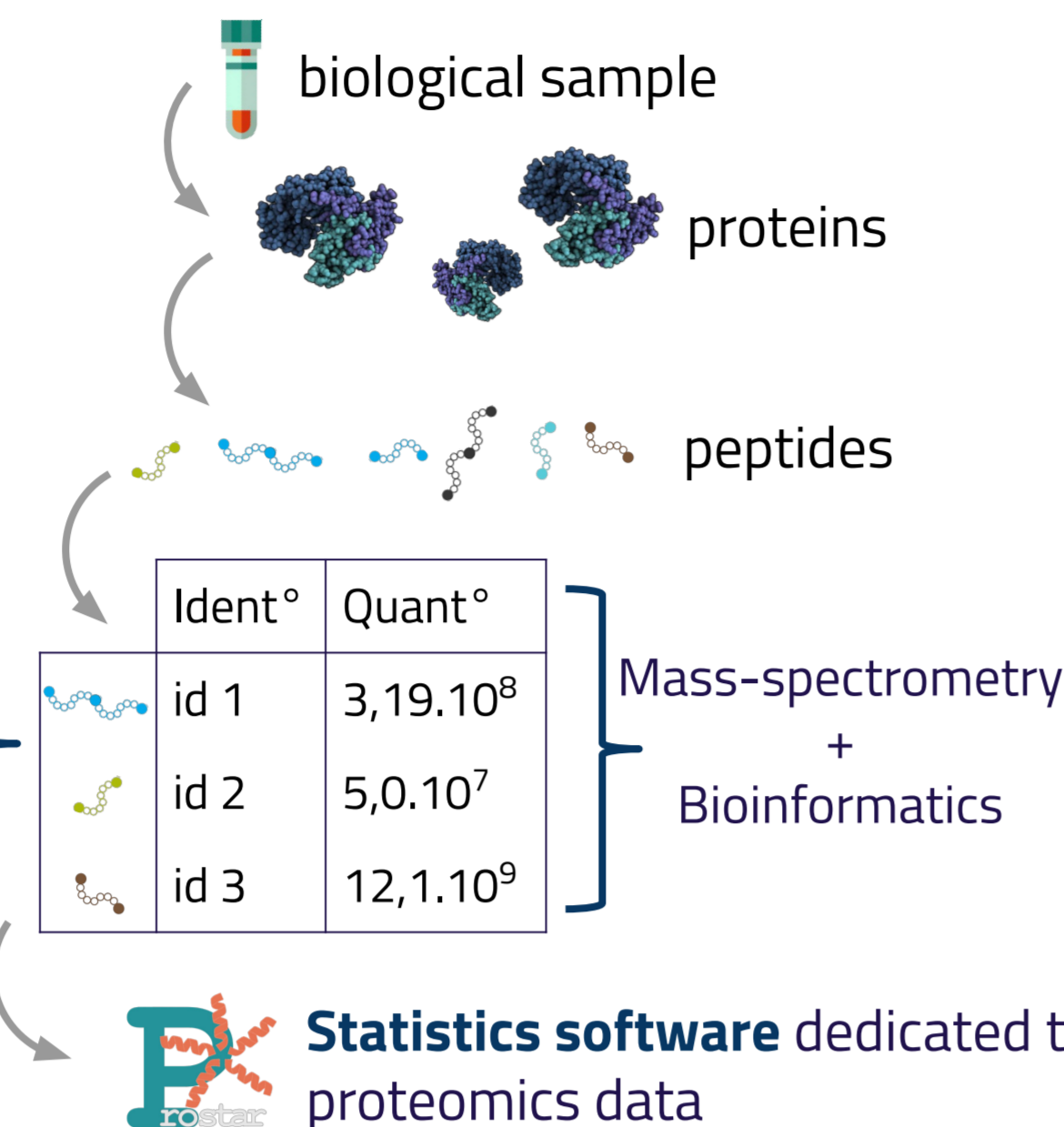
Quantitative Proteomics

GOAL: study proteins and reconstruct proteomes (sets of particular proteins corresponding to a given time and biological state).

STEPS:

- 1 Protein separation into peptides
- 2 identification
- 3 Quantification
- 4 Differential expression analysis

EDyP team



Issues

Pairwise comparisons of conditions only

BUT

Most biomedical studies have a complex experimental design

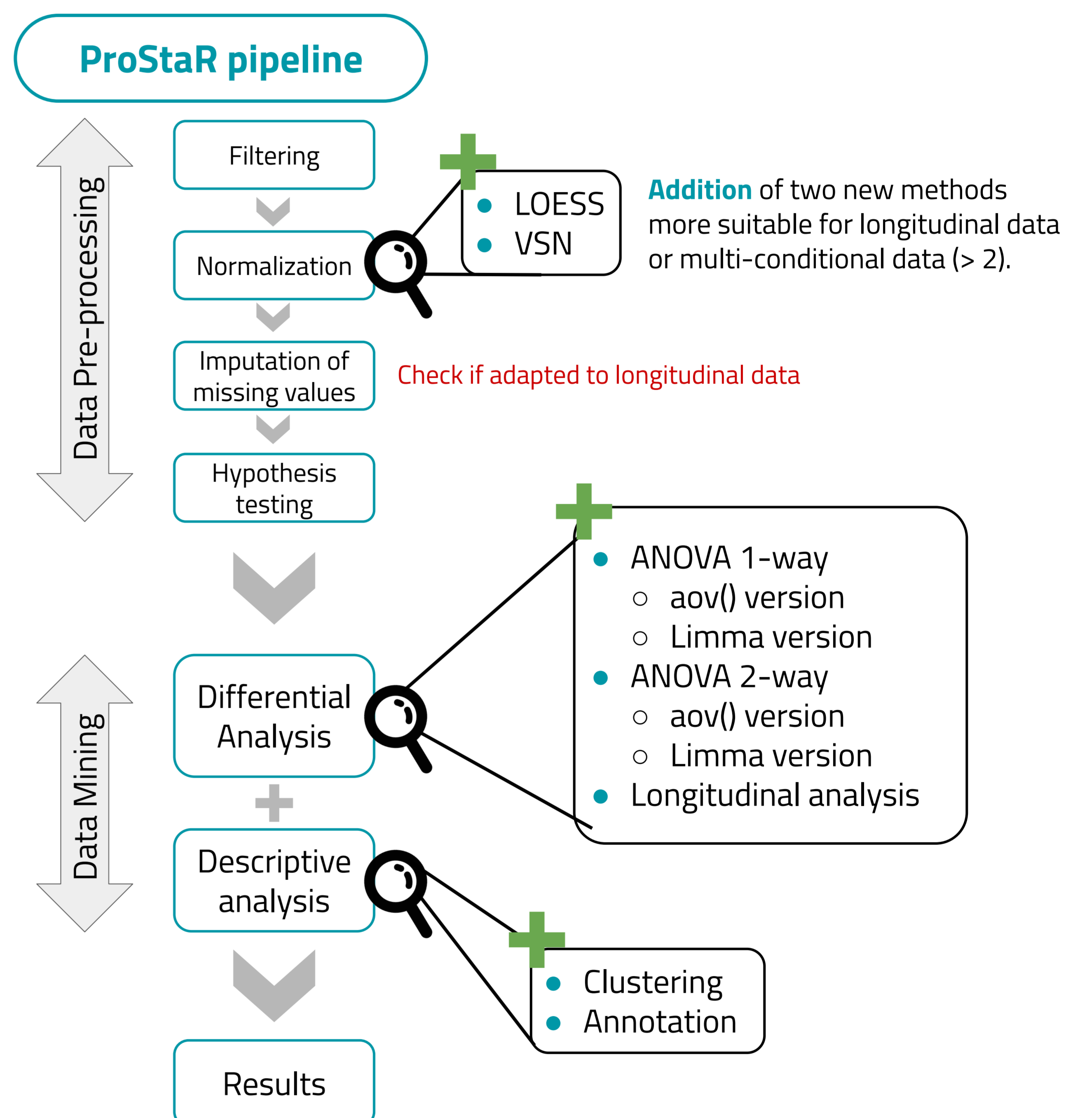
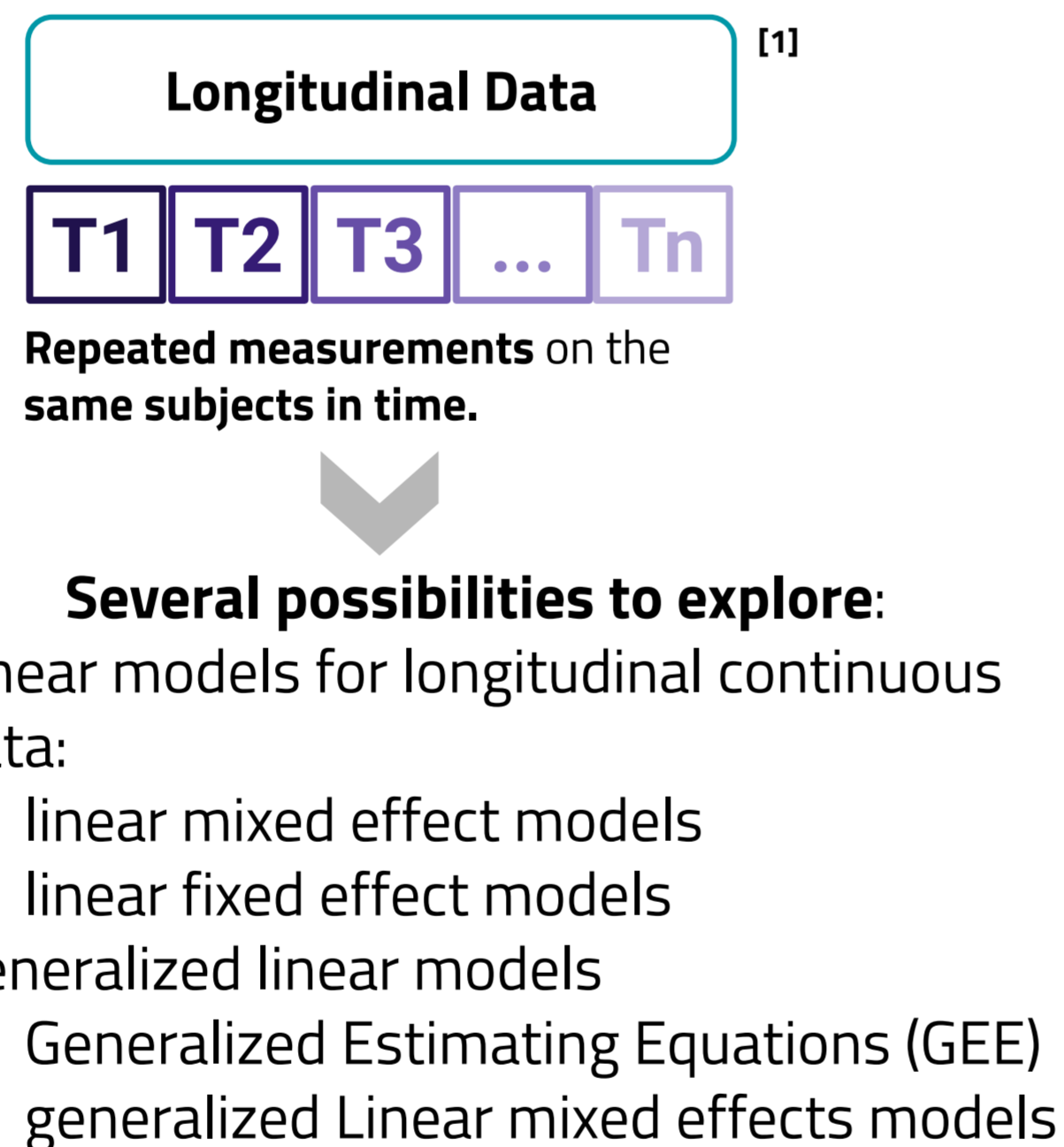
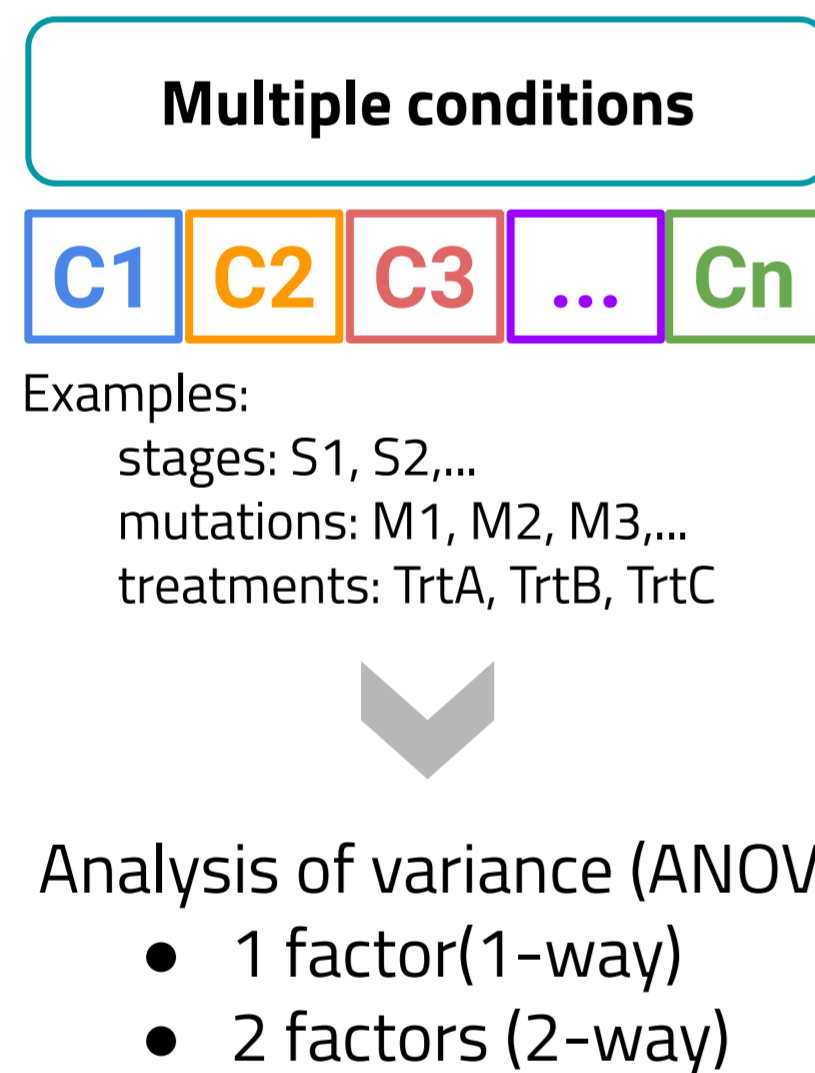
How to analyze them in a reliable & easy way?

2 Objectives of my thesis

1. Develop a rigorous statistical methodology adapted to proteomics data.
2. Allow the study of more complex experimental designs, such as longitudinal data, or with more than 2 experimental conditions to compare (multi-conditional data).
3. Improve the ProStaR software.
4. Use this method to look for potential biomarkers in some datasets.

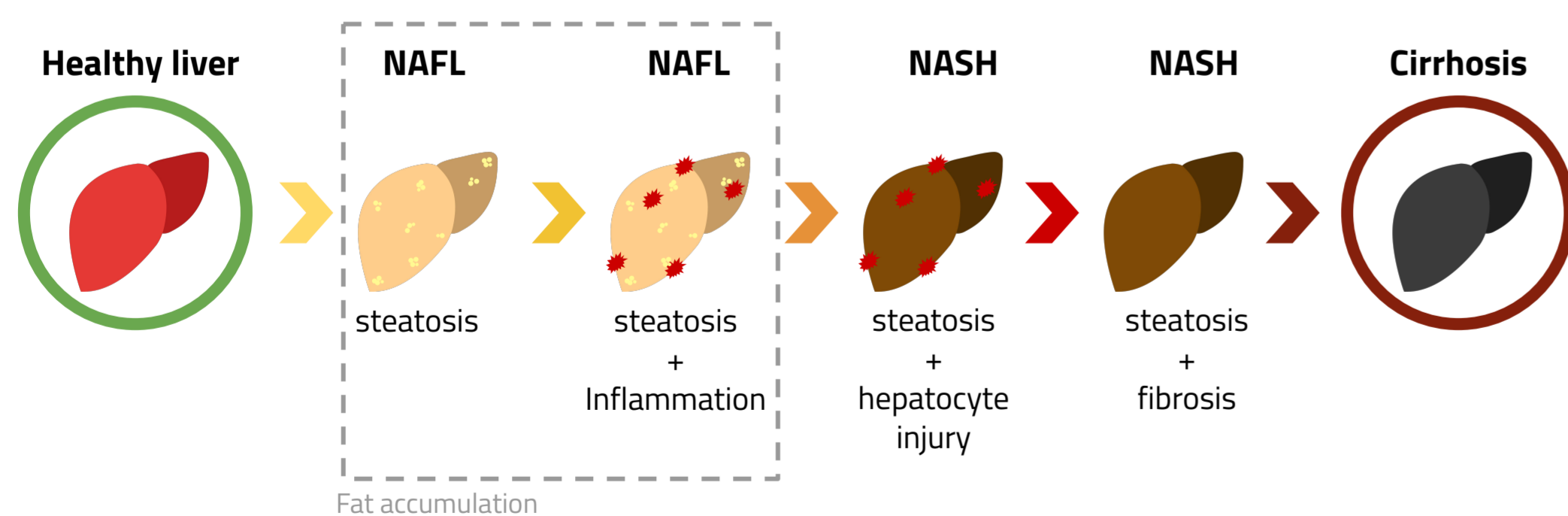
3 Methodology + Integration into ProStaR pipeline

1. Identification of statistical methods that will be adapted to different types of experimental designs at:
 - a. data normalization step
 - b. differential analysis step
2. Test the performance of these methods with simulated data by comparing them to more conventional methods (True Positive Rate, ...). Example: limma vs aov
3. Development and integration of code in R (language in which ProStaR is developed)
4. Preservation of steps at the pre-processing level.
5. Addition of new statistical tests at the level of the differential analysis:



4 Case study: Non Alcoholic SteatoHepatitis (NASH)

It is a progressive disease affecting the liver, which develops in 5 successive stages.



Biopsy is the only reliable method to determine at what stage a patient is. BUT it is a heavy and risky operation

Replace it with a blood sample and find biomarkers using proteomics

LIFE Project
University Hospital of Grenoble

- Cohort of 160 patients
 - 127 biological parameters measured but our work focuses on one of these: the SAF score
- it categorizes patients according to 3 key parameters:

Steatosis
Activity
Fibrosis

Our goal is to be able to categorize the patients in the same way as the SAF score, by using identified and quantified biomarkers.

5 Conclusion

Interdisciplinary thesis work: bring together Medicine, Biology, Informatics and Statistics to answer a complex question.

Contribution to Proteomics in terms of:

- data processing
- software engineering

Contribution to biomedical research

- Search for new potential biomarkers now (NASH LIFE project,...) and in the future.

Publications

- Wiecek S., Combes F., Borges H., Burger T. (2019) Protein-Level Statistical Analysis of Quantitative Label-Free Proteomics Data with ProStaR. In: Brun V., Couté Y. (eds) Proteomics for Biomarker Discovery. Methods in Molecular Biology, vol 1959. Humana Press, New York, NY
- Borges, H., Guibert, R., Permiakova, O., & Burger, T. (2018). Distinguishing between spectral clustering and cluster analysis of mass spectra. Journal of proteome research, 18(1), 571-573.

[1] Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2012). Applied longitudinal analysis (Vol. 998). John Wiley & Sons.

