POMA

Statistical analysis tool for targeted metabolomic data

Pol Castellano-Escuder

Jul 11, 2019 | useR! 2019 | Toulouse

Outline

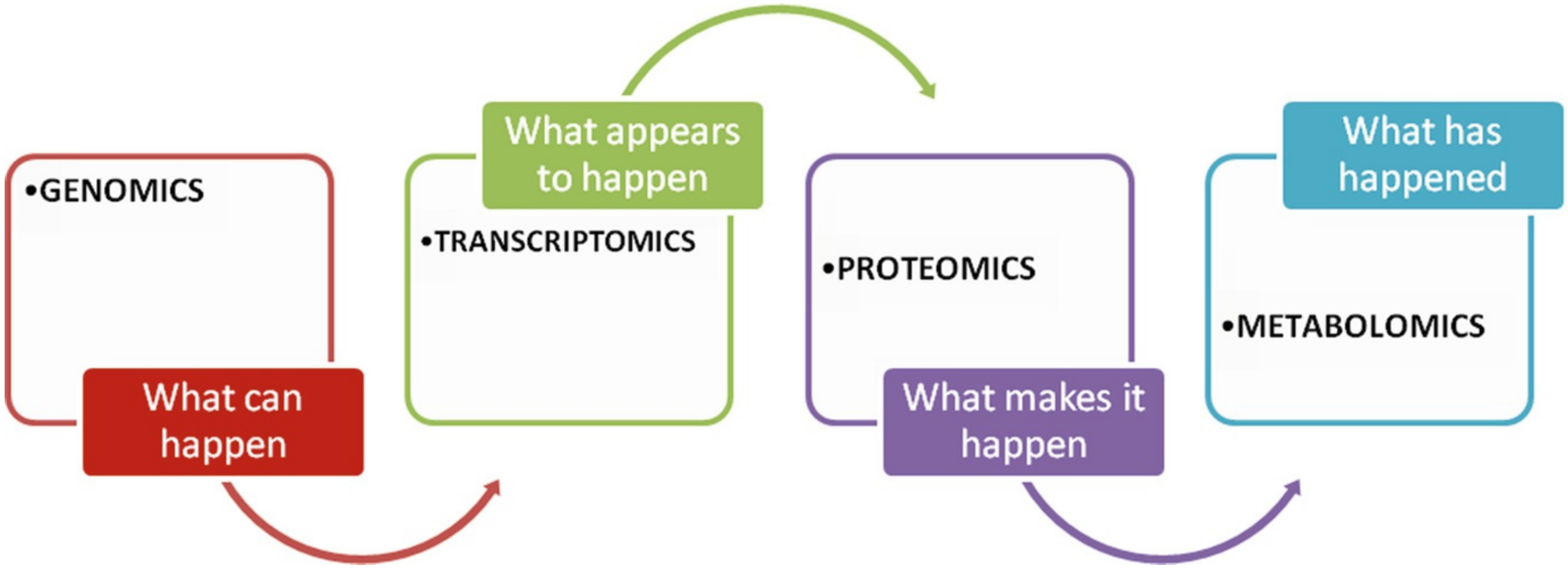
1. Context
2. Motivation & Aims
3. Results
4. Conclusions
5. Future Work

CONTEXT

What’s Metabolomics?

"Metabolomics is the identification and quantification of the small molecule metabolic products (the metabolome) of a biological system. Mass spectrometry and NMR spectroscopy are the techniques most often used for metabolome profiling"[1]

"The Omics Cascade"



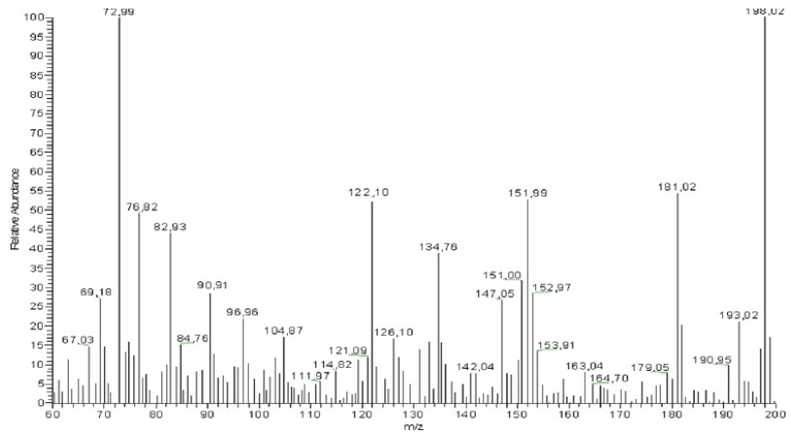
[2]

[1] https://www.nature.com/subjects/metabolomics

[2] Narad P., Kirthanashri S.V. (2018) Introduction to Omics. In: Arivaradarajan P., Misra G. (eds) Omics Approaches, Technologies And Applications. Springer, Singapore

The data

Targeted and untargeted metabolomics



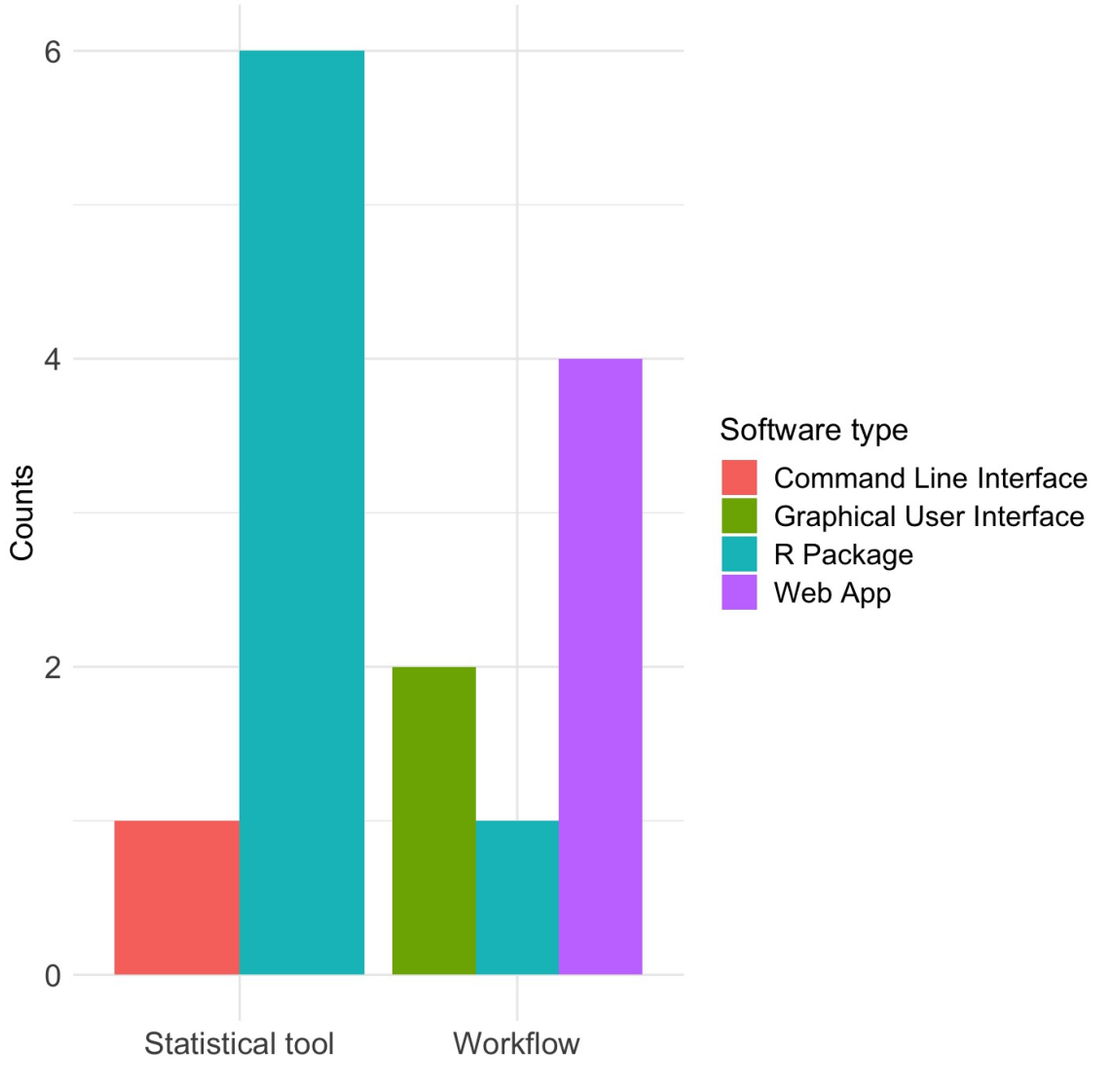
- **Targeted metabolomics:** we know the mass of the metabolites that we want to quantify BEFORE the analysis (hundreds)

- **Untargeted metabolomics:** all metabolites will be acquired, but we will not know exactly which ones are some of them (thousands)

How is the data that we will analize?

- Standard (Omics) matrix: Samples in rows and metabolites (variables) in columns

Freely Available Existing Tools



Web Apps that allows users to perform a statistical analysis[3]

- Workflow4metabolomics
- Galaxy-M
- XCMS Online
- MetaboAnalyst

[3] Spicer, R., Salek, R. M., Moreno, P., Cañueto, D., & Steinbeck, C. (2017). Navigating freely-available software tools for metabolomics analysis. Metabolomics, 13(9), 106.

MOTIVATION & AIMS

Motivation & Aims

Motivation

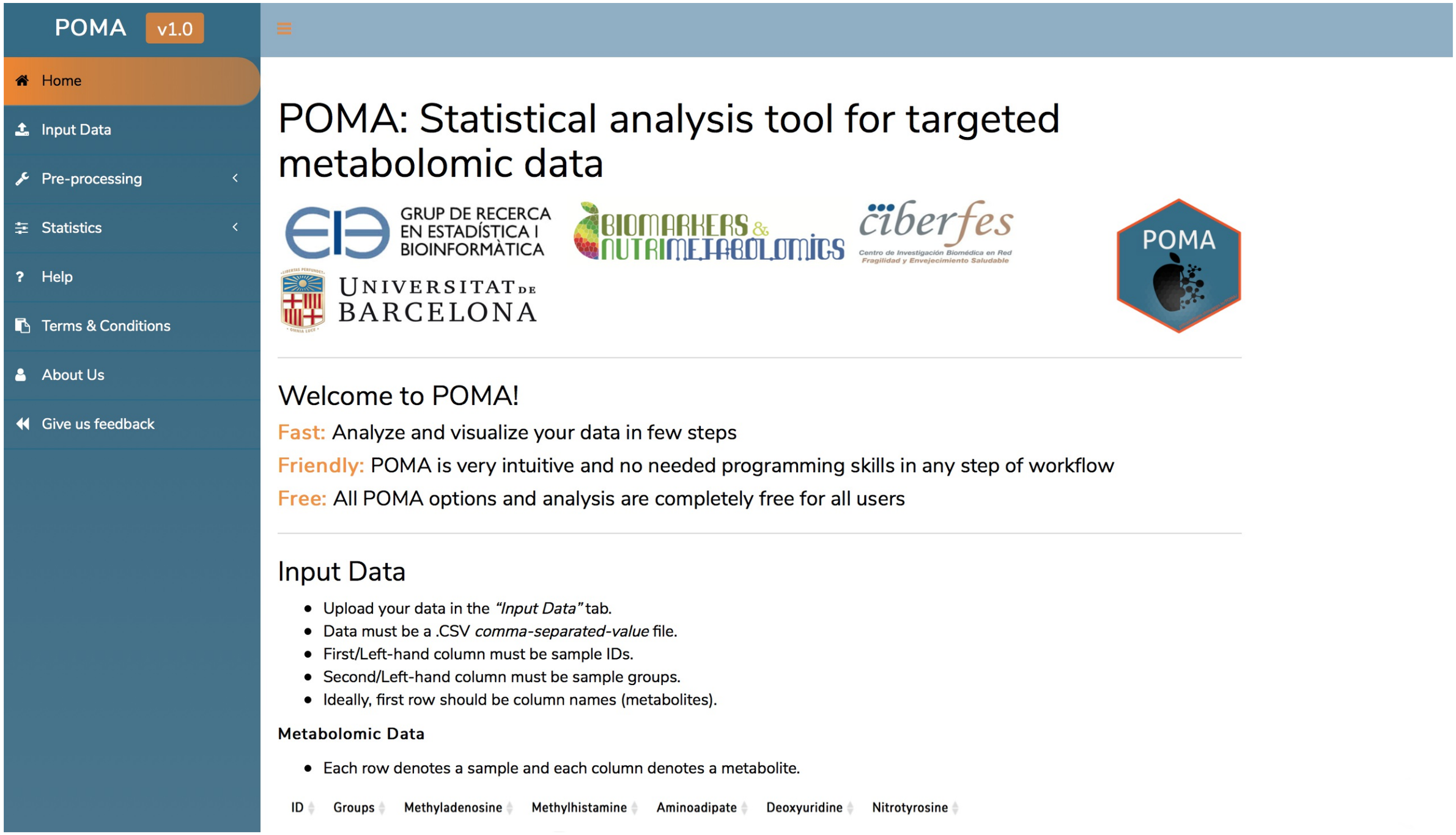
- Biological interpretation of the results is one of the hard points and high knowledge of statistical analysis and computational programming is usually required
- Sometimes, the existing tools don’t accept "complicated" databases

Aims

- Provide users of an **EASY USE** tool that don’t require programming skills
- Allow users to analyze all types of data (simple and complex)
- Lead the user for a good statistical analysis (Documentation & automatic reports)
- Make a completely **REPRODUCIBLE** analysis (Open Source)
- **Our main aim is COMPLETING the existing tools and give other option to users, NOT to COMPETE with the existing tools**

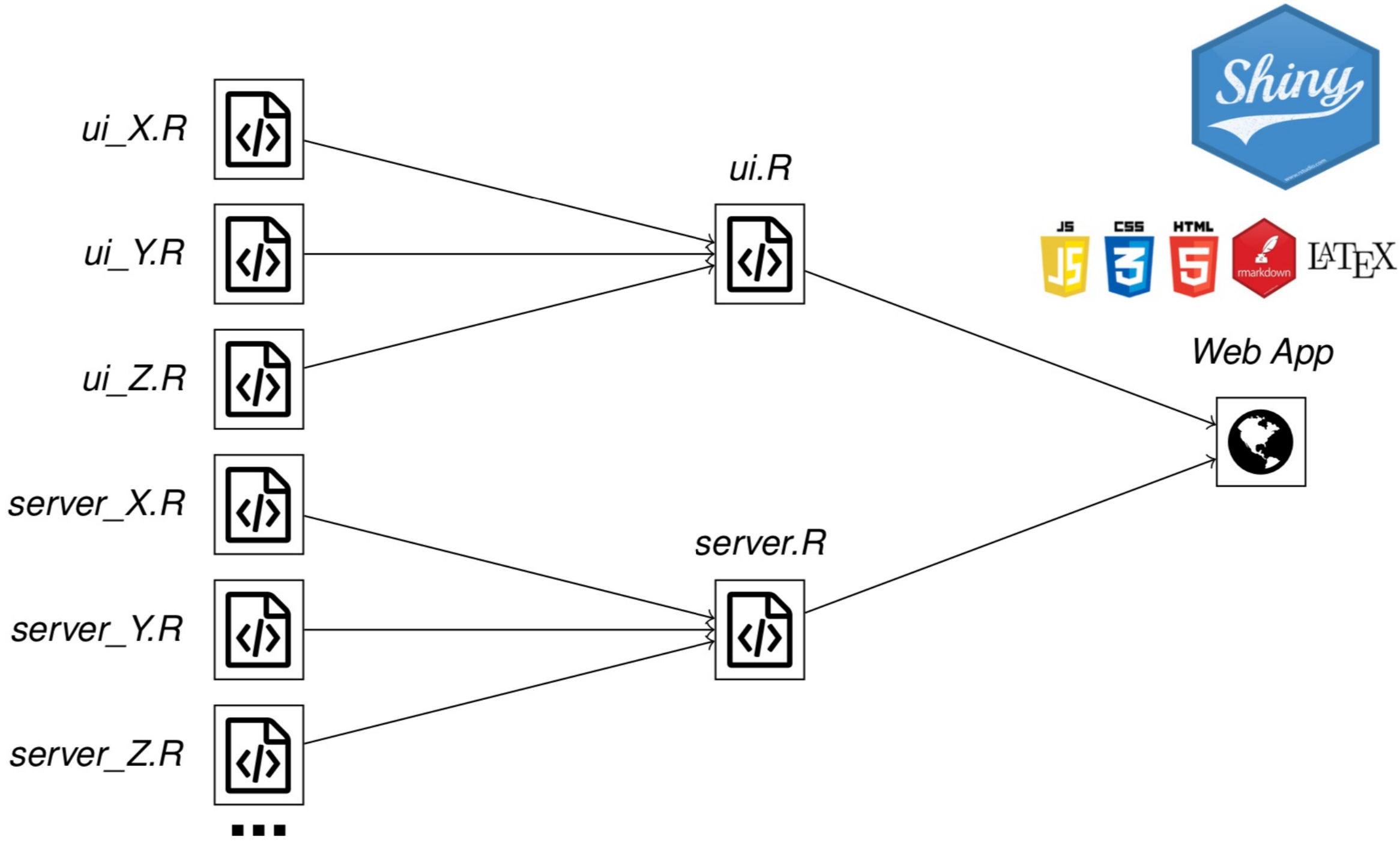
RESULTS

POMA Shiny App



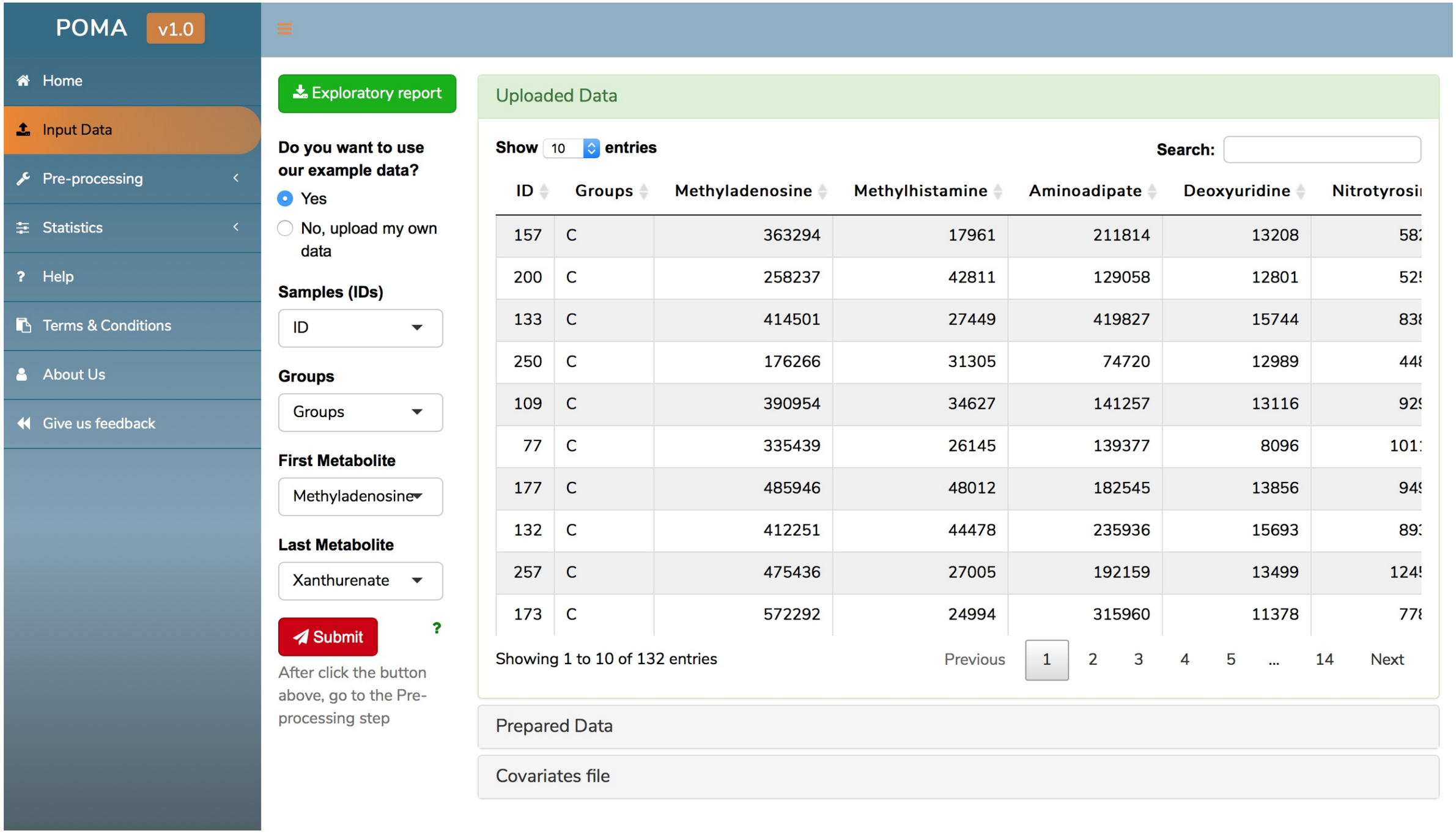
http://polcastellano.shinyapps.io/POMA/

Architecture



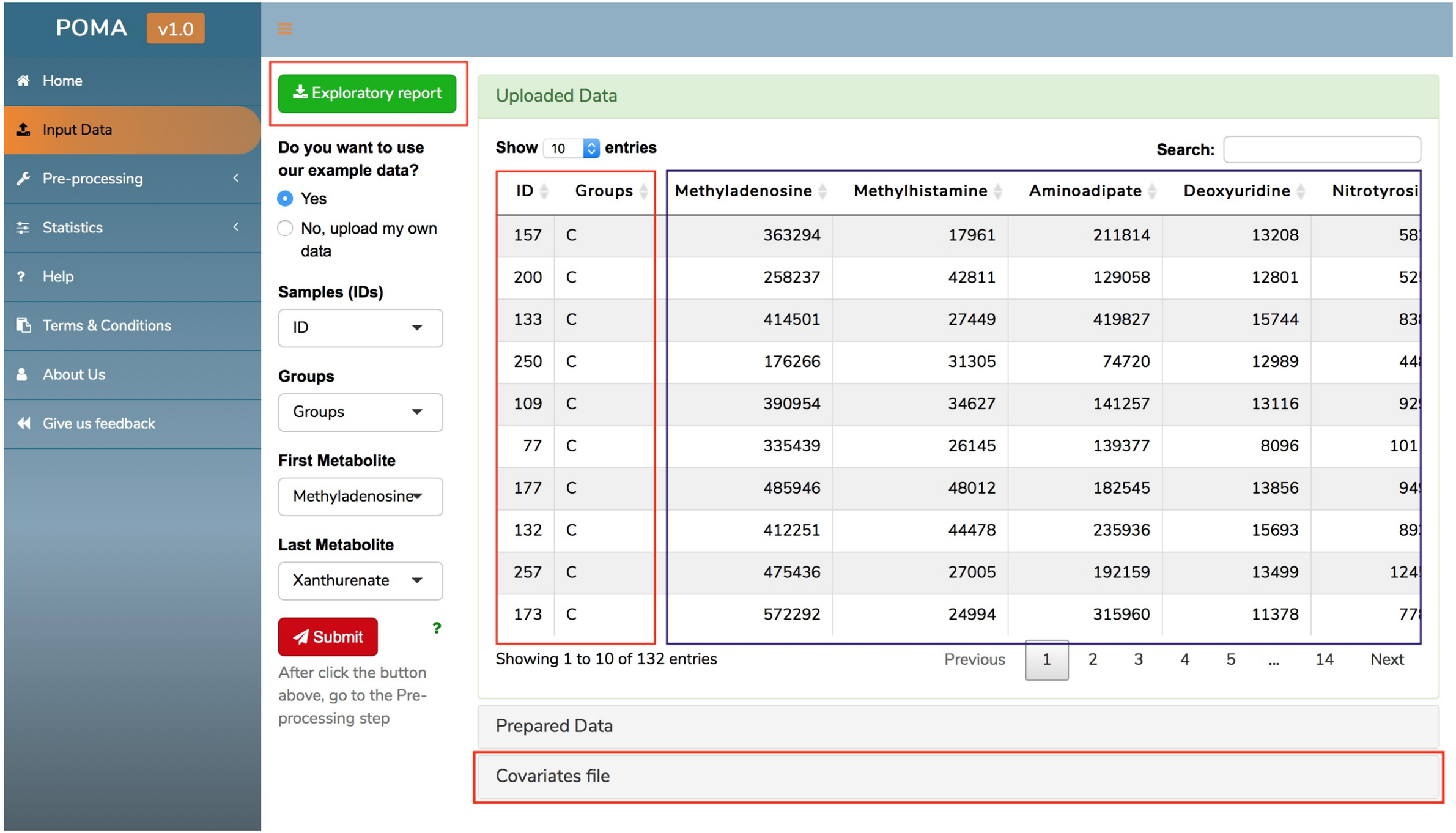
Input Data Panel

We have used the `shinydashboard` package for the main structure and the `dashboardthemes` package for customization



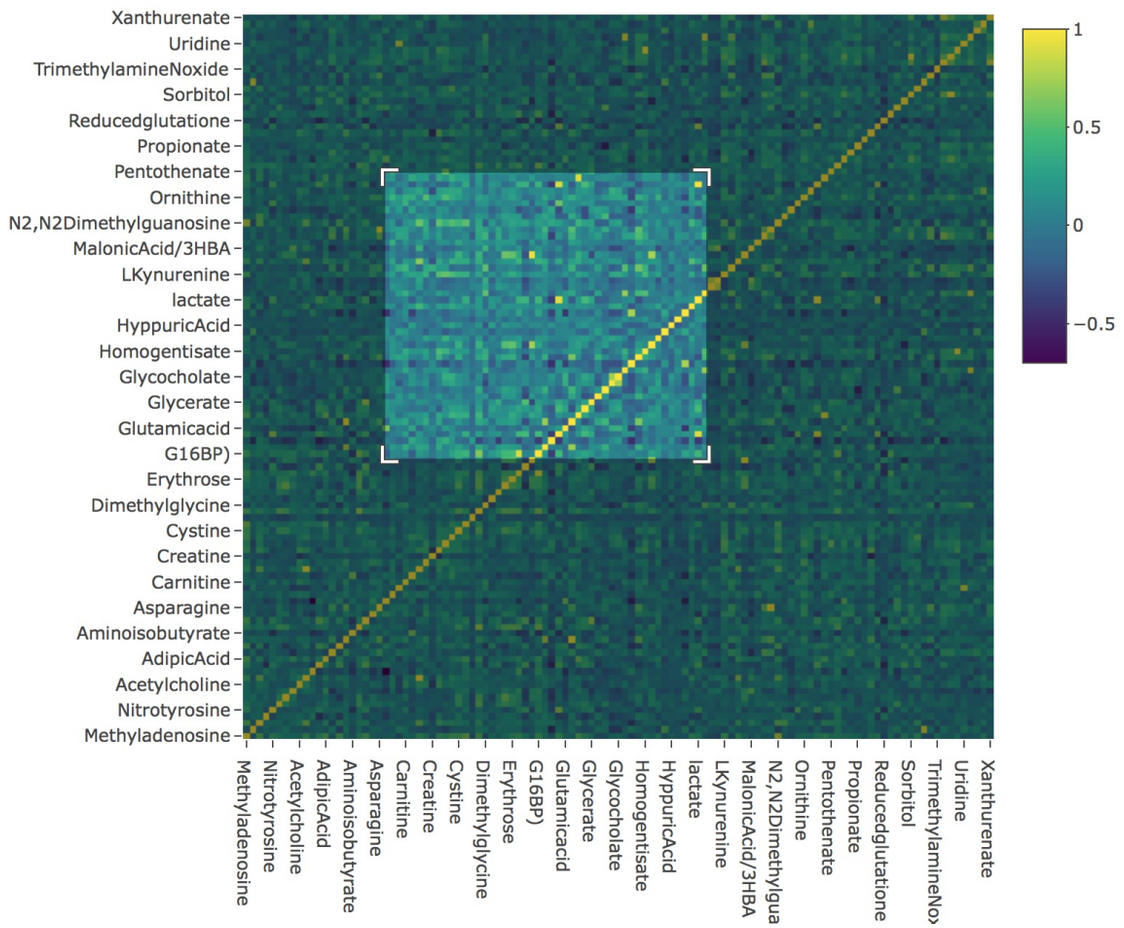
Input Data Panel

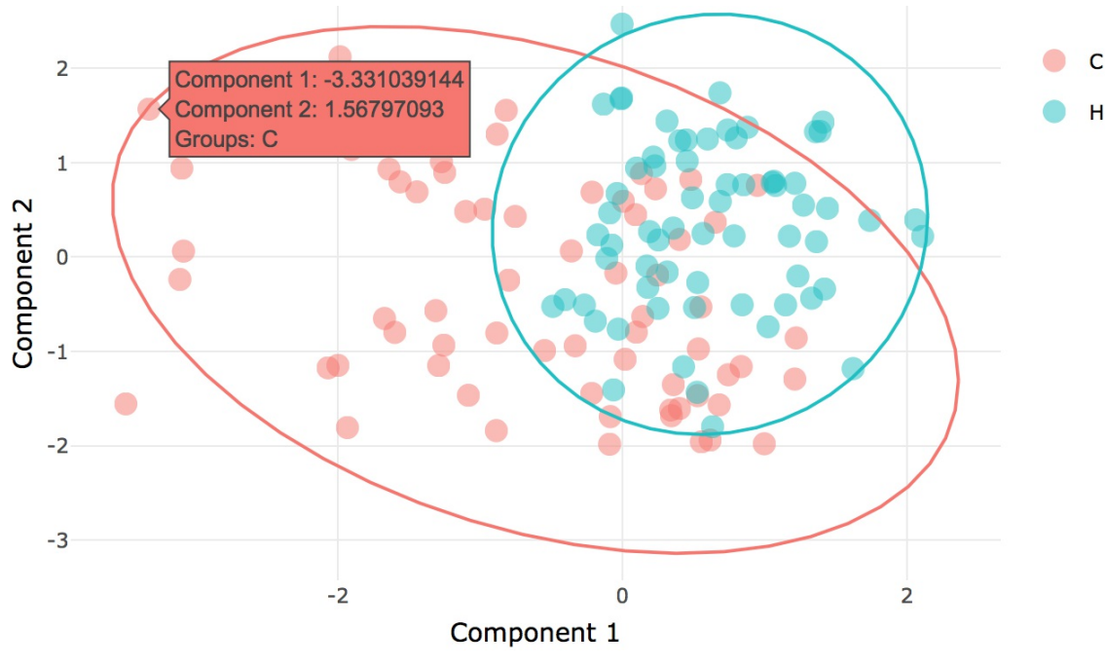
We have used the `shinydashboard` package for the main structure and the `dashboardthemes` package for customization



Visualization

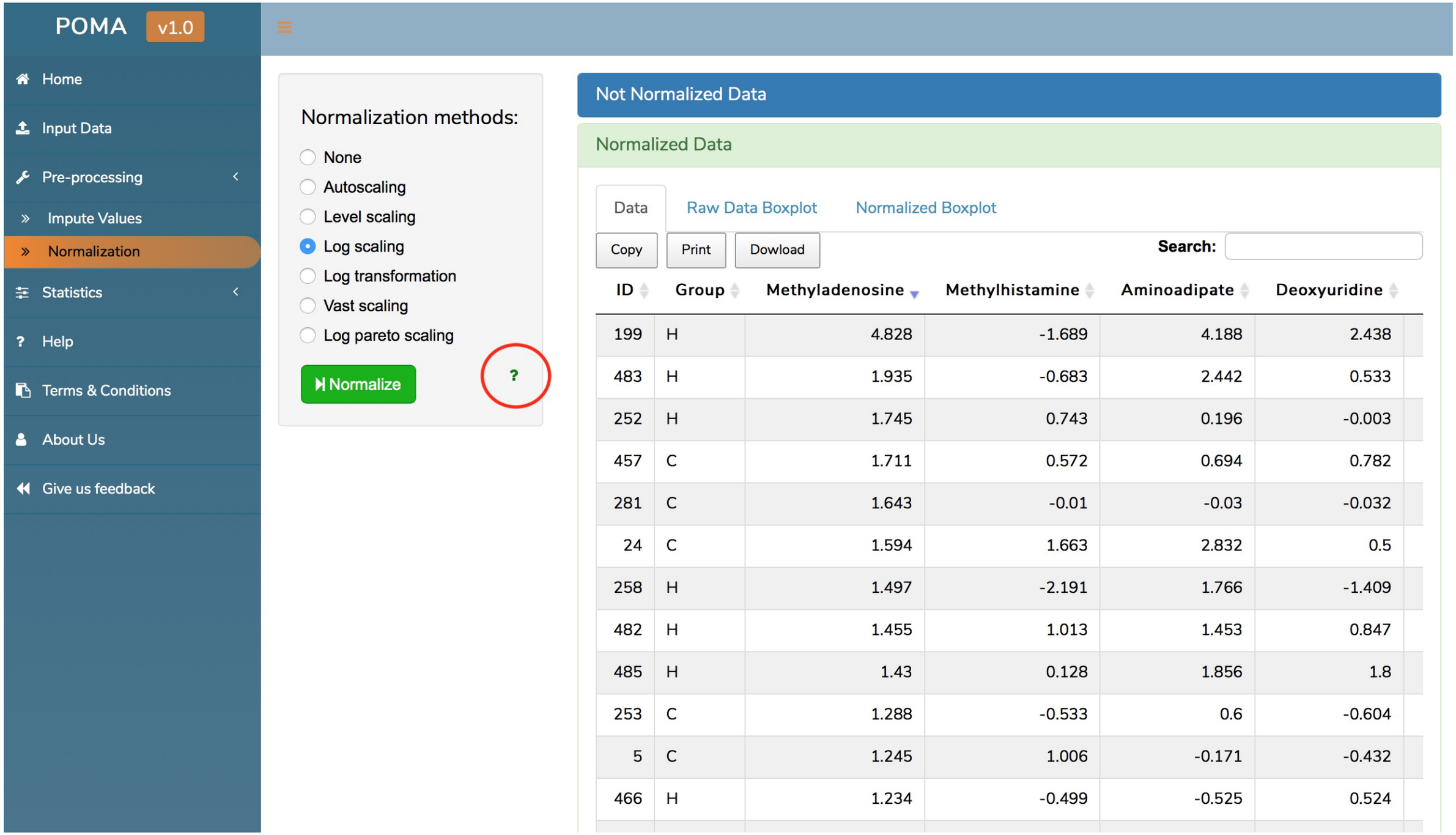
All plots in the app are designed using `plotly` package. It make all plots interactive allowing users to zoom in or zoom out in a plots, select points to see the individual information, hide all points of one group and download plots in a easy way!





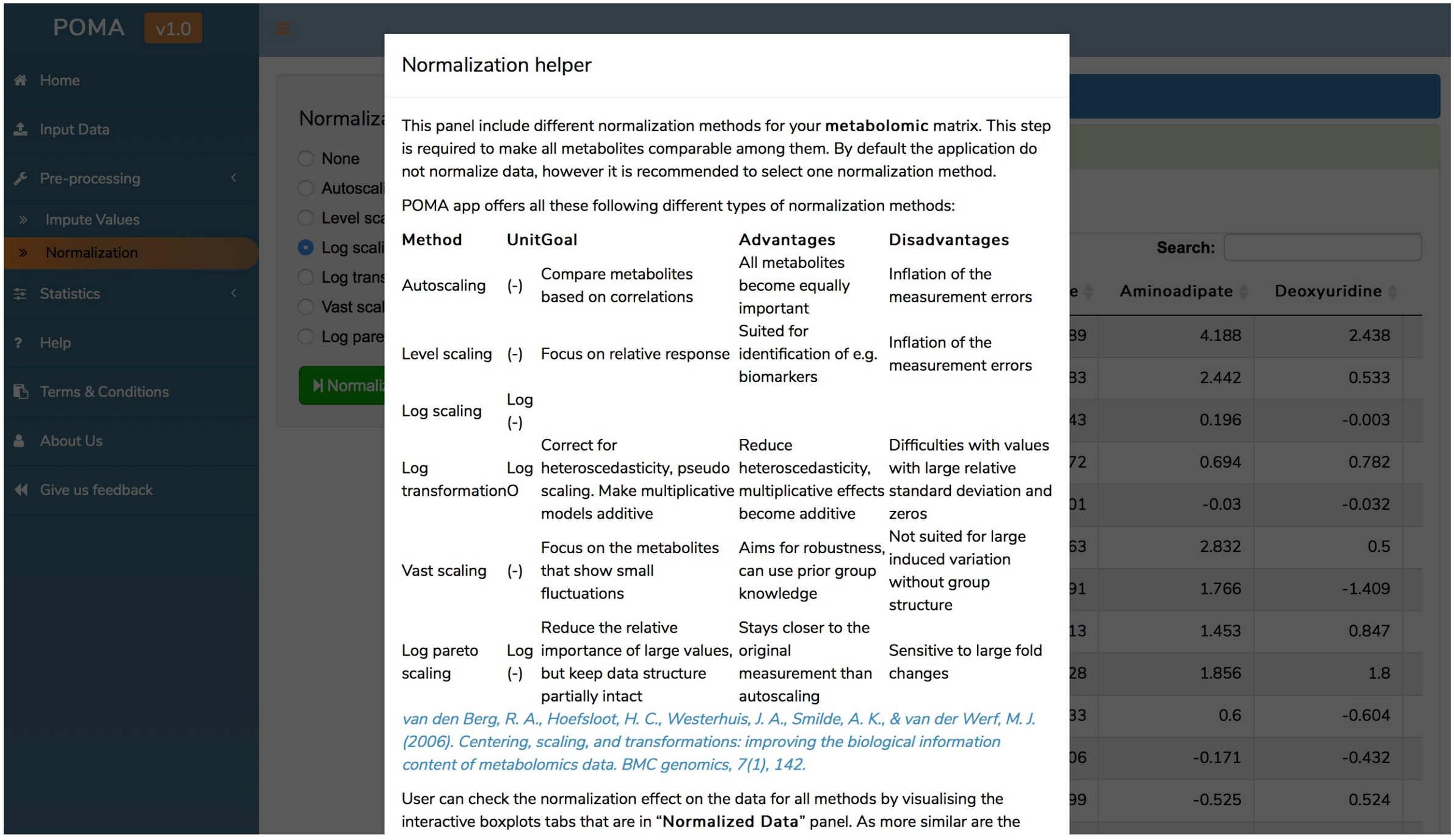
Documentation

The implementation of `shinyhelper` package allows each panel to have an individual help



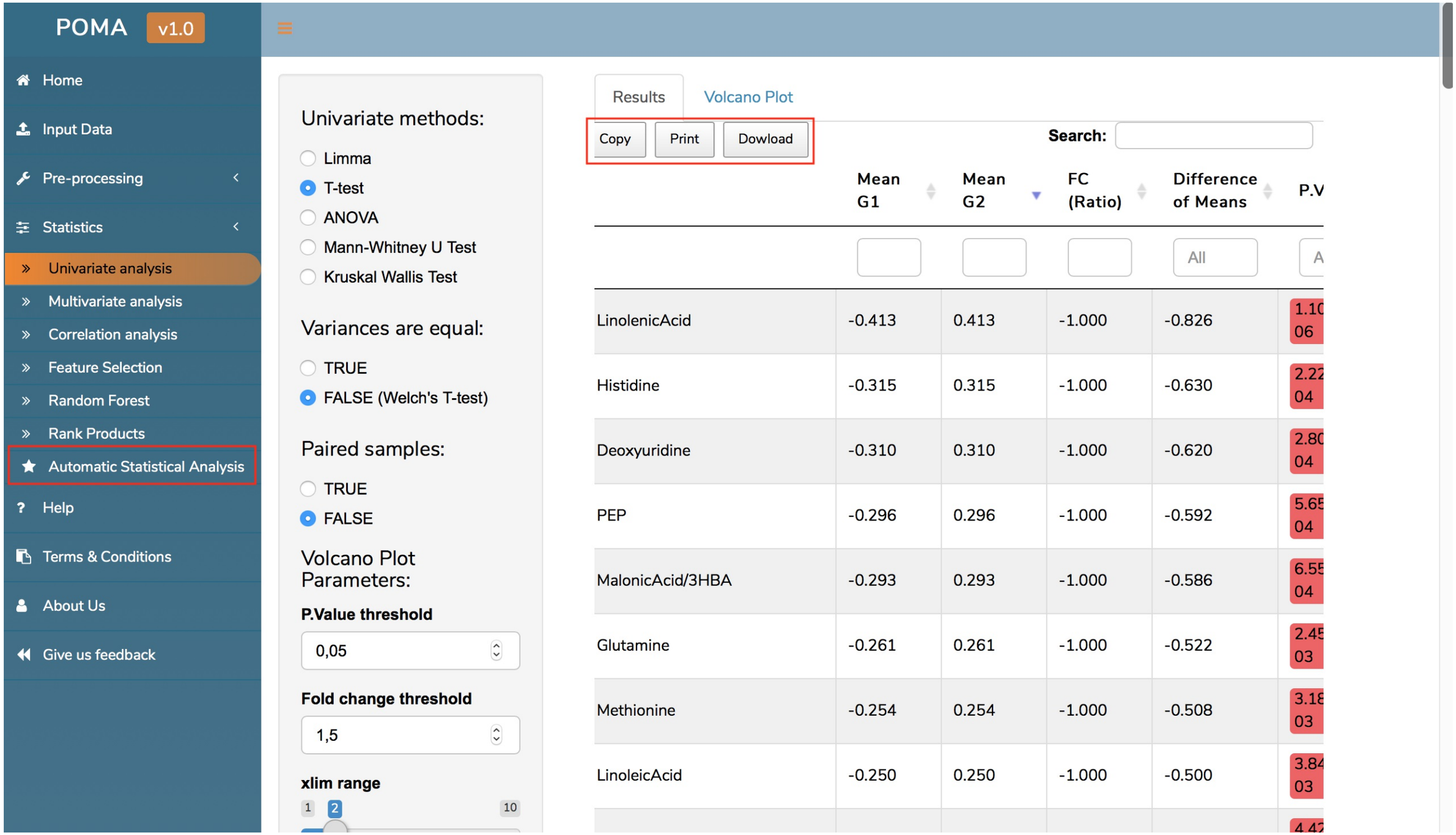
Documentation

The implementation of `shinyhelper` package allows each panel to have an individual help



Statistical Analysis

The aim is to offer to tune as many parameters as possible to avoid the "black box" effect



Automatic Statistical Report

POMA: Statistical analysis tool for targeted metabolomic data

Intelligent Statistical Analysis: Metabolomic analysis for 2 groups using default 'Pre-processing' by POMA

July, 2019

- 1 Parametric tests
  - 1.1 T-test
    - 1.1.1 Metabolites with NORMAL distribution & variance HOMOSCEDASTICITY
    - 1.1.2 Metabolites with NORMAL distribution & variance HETEROSCEDASTICITY
  - 1.2 ANOVA
    - 1.2.1 Metabolites with NORMAL distribution ANOVA model
- 2 Non Parametric tests
  - 2.1 Mann-Whitney U Test (Wilcoxon Signed Rank Test if the data is paired)
    - 2.1.1 Metabolites with NON NORMAL distribution

1 Parametric tests

1.1 T-test

1.1.1 Metabolites with NORMAL distribution & HOMOSCEDASTICITY

Metabolite	Mean G1	Mean G2	FC (Ratio)	Difference of Means	P.Value	adj.P.Val
Deoxyuridine	4.050	4.097	1.012	-0.047	0.00028	0.01593
Glycochenodeoxycholate	5.526	5.183	0.938	0.343	0.00074	0.02104
MaleicAcid)	6.297	6.240	0.991	0.057	0.00114	0.02173
Methionine	5.782	5.832	1.009	-0.050	0.00315	0.04493
Allantoin	5.034	4.921	0.978	0.113	0.00650	0.07409

CONCLUSIONS

Conclusions

- **We have developed a FAST, FRIENDLY and FREE software that is called POMA**
- **POMA is full-based in R language and uses a Shiny system to run**
- **POMA provides an accurate DOCUMENTATION ("HELP") at each step of analysis that coluld improve the results and facilitate the interpretation of it**
- **POMA can generate two types of AUTOMATIC REPORTS: Exploratory report and Statistical report**
- **POMA is in a constant development. According to this, we are totally open to user bug reports to keep improving our app**

Future Work

(In order of importance...)

- Finishing the **documentation** as accurately as possible
- Make the code more **efficient**
- Develop a **package** with all POMA functions

Thank you all!

To the Statistics and Bioinformatics Research Group and Biomarkers and Nutritional & Food Metabolomics Research Group from University of Barcelona for amazing support

To the useR! 2019 organizers, for allowing me to show this work

Slides created via the R package xaringan

---

polcaes@gmail.com
pcastellanoescuder.github.io/
@polcastellano_
@pcastellanoescuder
University of Barcelona