



# 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



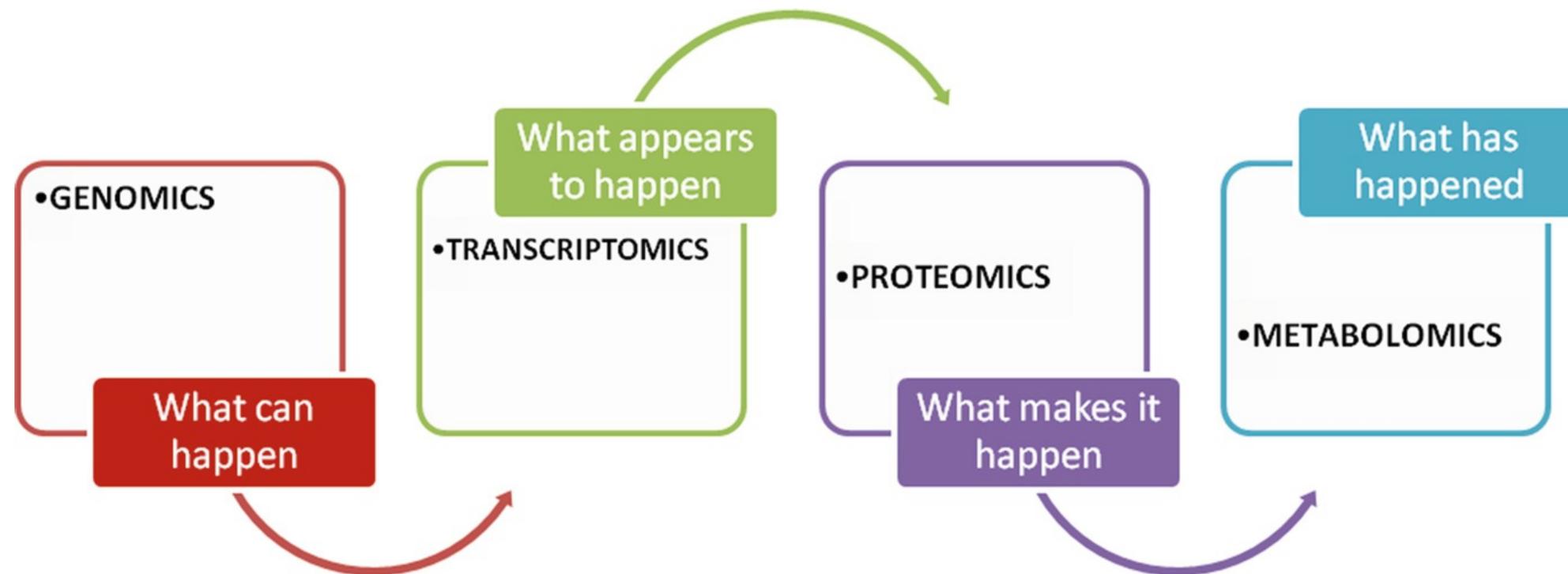
# CONTEXT

---

# 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"<sup>1</sup>

## "The Omics Cascade"



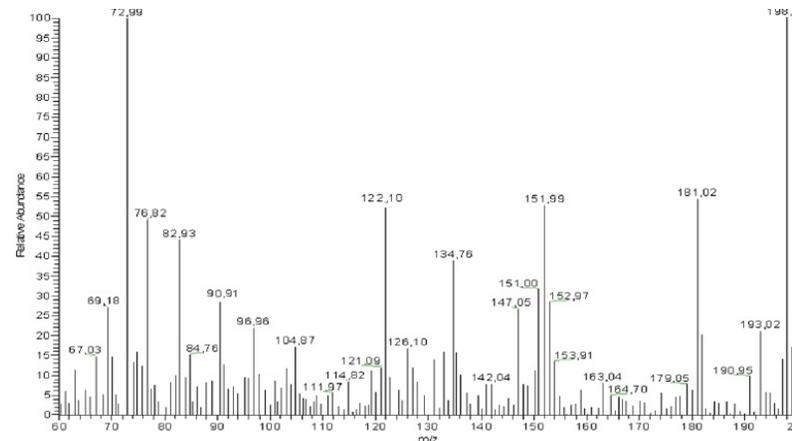
2

[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

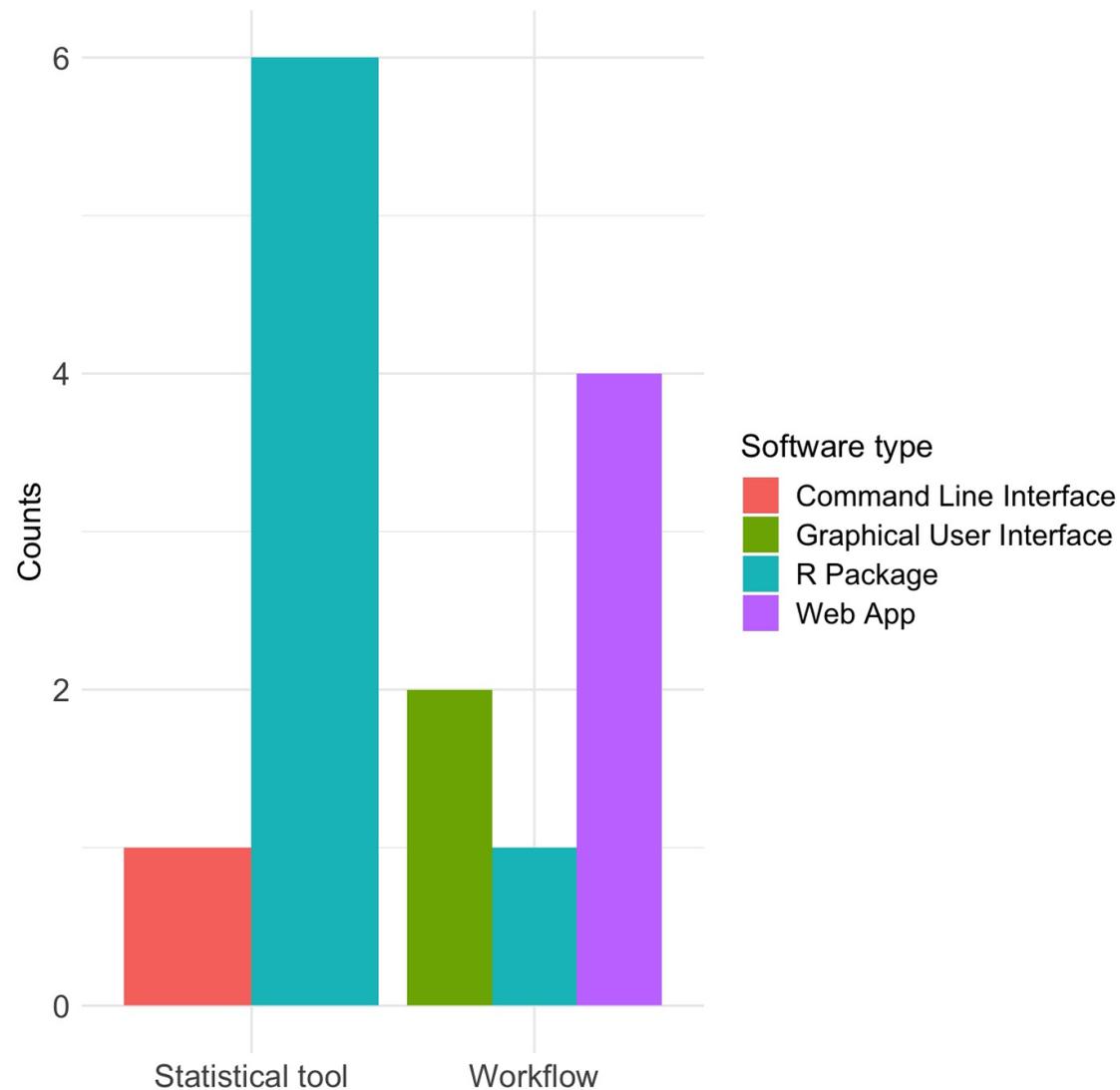


- **Targeted metabolomics:** we know the mass of the metabolites that we want to quantify BEFORE the analysis (hundreds)
- **Untargeted metabolomics:** all metabolites will be acquired, but we will not know exactly which ones are some of them (thousands)

## How is the data that we will analyze?

- 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<sup>3</sup>

- 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

POMA v1.0

Home

Input Data

Pre-processing

Statistics

Help

Terms & Conditions

About Us

Give us feedback

## POMA: Statistical analysis tool for targeted metabolomic data

 GRUP DE RECERCA EN ESTADÍSTICA I BIOINFORMÀTICA

 BIOMARKERS & NUTRIMETABOLOMICS

 ciberfes  
Centro de Investigación Biomédica en Red  
Fragilidad y Envejecimiento Saludable

 UNIVERSITAT DE BARCELONA



---

### Welcome to POMA!

**Fast:** Analyze and visualize your data in few steps

**Friendly:** POMA is very intuitive and no needed programming skills in any step of workflow

**Free:** All POMA options and analysis are completely free for all users

---

### Input Data

- Upload your data in the "Input Data" tab.
- Data must be a .CSV *comma-separated-value* file.
- First/Left-hand column must be sample IDs.
- Second/Left-hand column must be sample groups.
- Ideally, first row should be column names (metabolites).

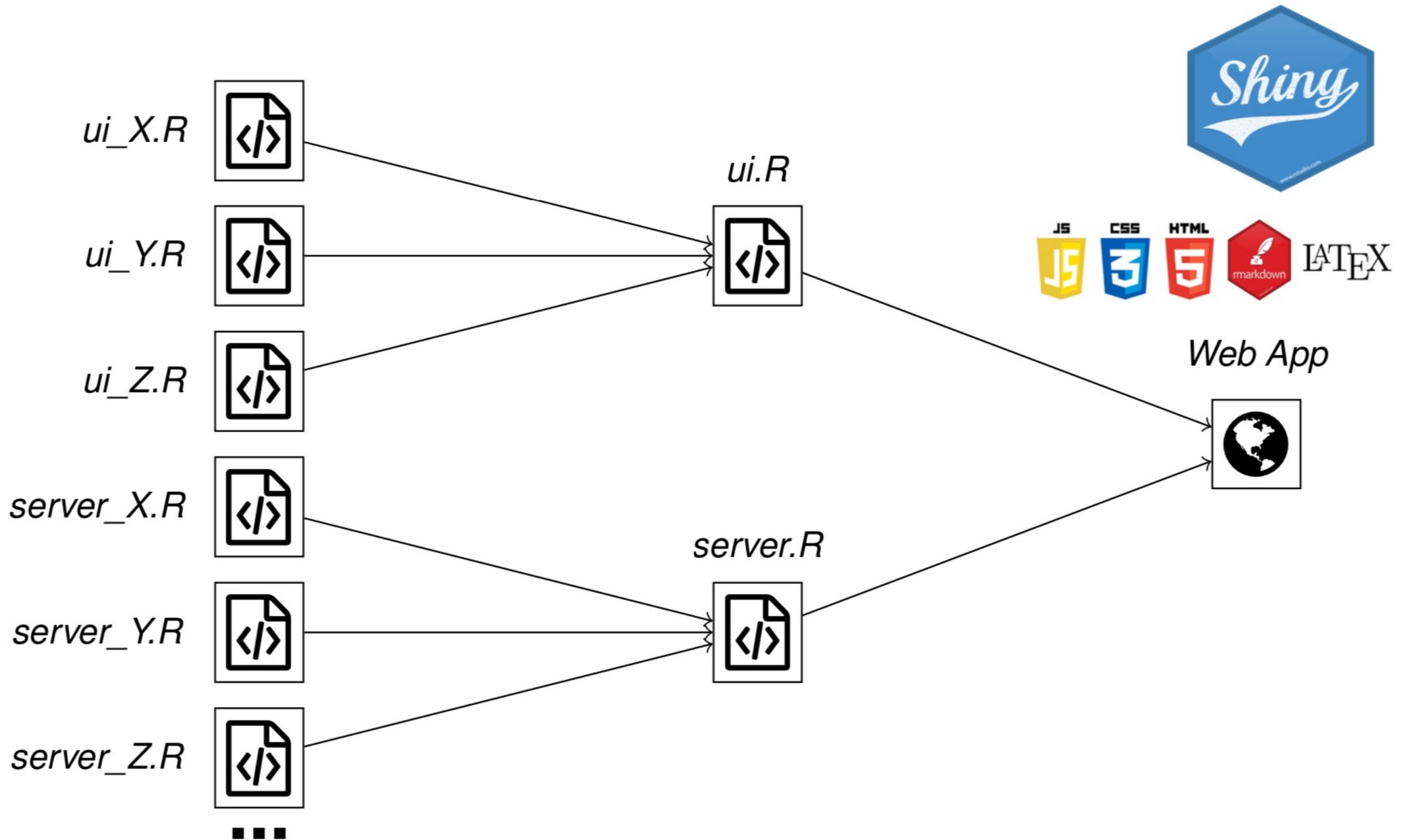
### Metabolomic Data

- Each row denotes a sample and each column denotes a metabolite.

ID ⌵ Groups ⌵ Methyladenosine ⌵ Methylhistamine ⌵ Amino adipate ⌵ Deoxyuridine ⌵ Nitrotyrosine ⌵

<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

**POMA v1.0**

- Home
- Input Data**
- Pre-processing
- Statistics
- Help
- Terms & Conditions
- About Us
- Give us feedback

**Submitted Data**

Do you want to use our example data?  
 Yes  
 No, upload my own data

**Samples (IDs)**  
ID

**Groups**  
Groups

**First Metabolite**  
Methyladenosine

**Last Metabolite**  
Xanthurenate

**Submit** ?  
After click the button above, go to the Pre-processing step

**Submitted Data**

Show 10 entries Search:

ID	Groups	Methyladenosine	Methylhistamine	Amino adipate	Deoxyuridine	Nitrotyrosine
157	C	363294	17961	211814	13208	588
200	C	258237	42811	129058	12801	528
133	C	414501	27449	419827	15744	838
250	C	176266	31305	74720	12989	448
109	C	390954	34627	141257	13116	928
77	C	335439	26145	139377	8096	1018
177	C	485946	48012	182545	13856	948
132	C	412251	44478	235936	15693	898
257	C	475436	27005	192159	13499	1248
173	C	572292	24994	315960	11378	778

Showing 1 to 10 of 132 entries Previous 1 2 3 4 5 ... 14 Next

**Prepared Data**

**Covariates file**

# Input Data Panel

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

**POMA v1.0**

Home

**Input Data**

Pre-processing

Statistics

Help

Terms & Conditions

About Us

Give us feedback

**Exploratory report**

Do you want to use our example data?

Yes

No, upload my own data

**Samples (IDs)**

ID

**Groups**

Groups

**First Metabolite**

Methyladenosine

**Last Metabolite**

Xanthurenate

**Submit**

After click the button above, go to the Pre-processing step

**Uploaded Data**

Show 10 entries Search:

ID	Groups	Methyladenosine	Methylhistamine	Amino adipate	Deoxyuridine	Nitrotyrosine
157	C	363294	17961	211814	13208	58
200	C	258237	42811	129058	12801	52
133	C	414501	27449	419827	15744	83
250	C	176266	31305	74720	12989	44
109	C	390954	34627	141257	13116	92
77	C	335439	26145	139377	8096	101
177	C	485946	48012	182545	13856	94
132	C	412251	44478	235936	15693	89
257	C	475436	27005	192159	13499	124
173	C	572292	24994	315960	11378	77

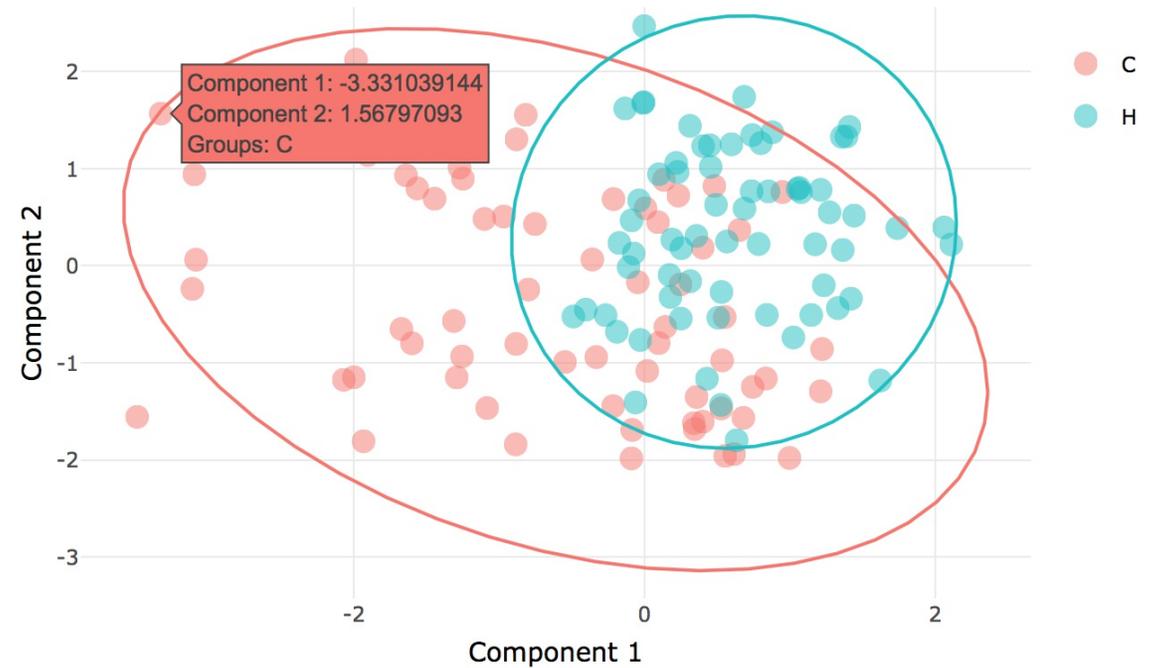
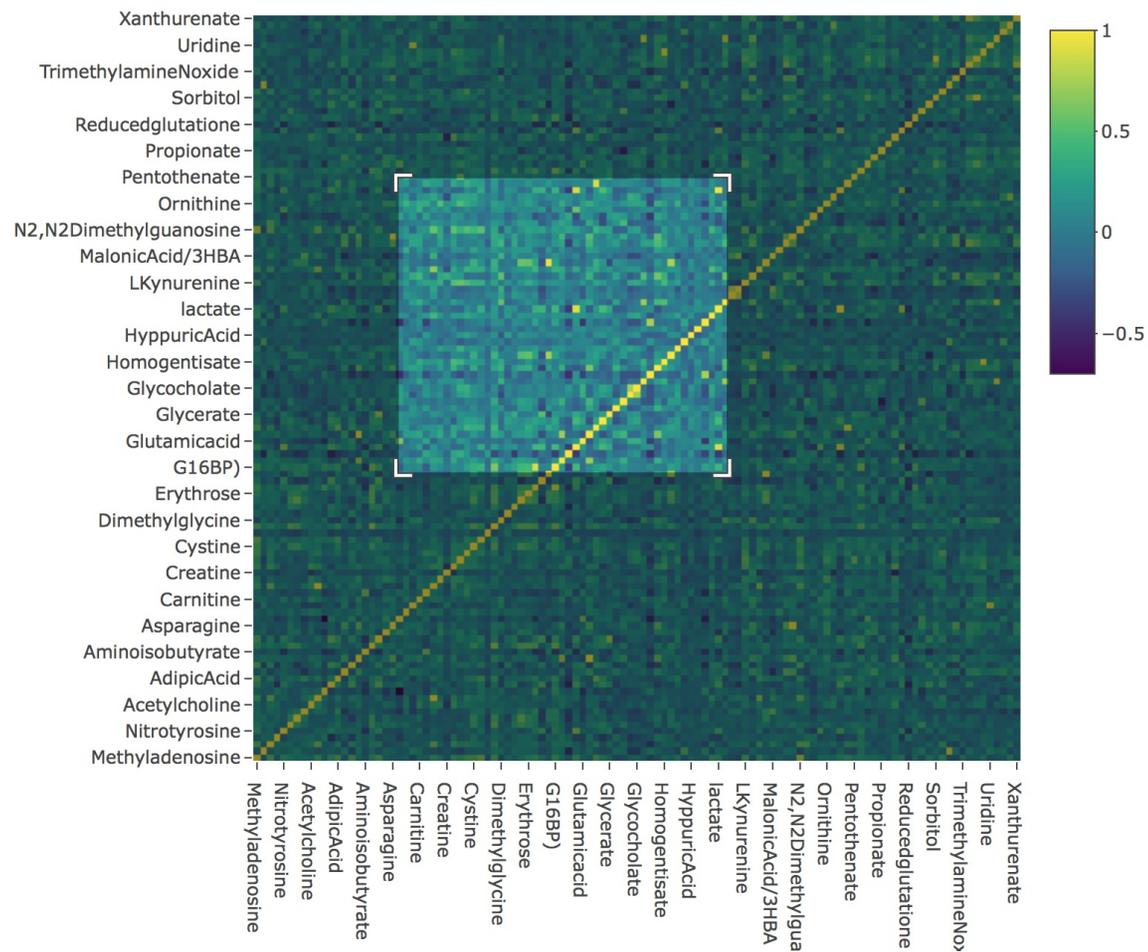
Showing 1 to 10 of 132 entries Previous 1 2 3 4 5 ... 14 Next

**Prepared Data**

Covariates file

# 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

The screenshot displays the POMA v1.0 web application interface. On the left is a navigation sidebar with options: Home, Input Data, Pre-processing, Impute Values, Normalization (highlighted), Statistics, Help, Terms & Conditions, About Us, and Give us feedback. The main content area is divided into two panels. The left panel, titled 'Normalization methods:', lists several options: None, Autoscaling, Level scaling, Log scaling (selected), Log transformation, Vast scaling, and Log pareto scaling. Below these is a green 'Normalize' button and a red circle containing a question mark icon. The right panel, titled 'Not Normalized Data', shows a 'Normalized Data' section with tabs for 'Data', 'Raw Data Boxplot', and 'Normalized Boxplot'. It includes buttons for 'Copy', 'Print', and 'Download', a search bar, and a table of data.

ID	Group	Methyladenosine	Methylhistamine	Aminoadipate	Deoxyuridine
199	H	4.828	-1.689	4.188	2.438
483	H	1.935	-0.683	2.442	0.533
252	H	1.745	0.743	0.196	-0.003
457	C	1.711	0.572	0.694	0.782
281	C	1.643	-0.01	-0.03	-0.032
24	C	1.594	1.663	2.832	0.5
258	H	1.497	-2.191	1.766	-1.409
482	H	1.455	1.013	1.453	0.847
485	H	1.43	0.128	1.856	1.8
253	C	1.288	-0.533	0.6	-0.604
5	C	1.245	1.006	-0.171	-0.432
466	H	1.234	-0.499	-0.525	0.524

# Documentation

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

**Normalization helper**

This panel include different normalization methods for your **metabolomic** matrix. This step is required to make all metabolites comparable among them. By default the application do not normalize data, however it is recommended to select one normalization method.

POMA app offers all these following different types of normalization methods:

Method	Unit	Goal	Advantages	Disadvantages
Autoscaling	(-)	Compare metabolites based on correlations	All metabolites become equally important Suited for identification of e.g. biomarkers	Inflation of the measurement errors
Level scaling	(-)	Focus on relative response		Inflation of the measurement errors
Log scaling	Log (-)	Correct for heteroscedasticity, pseudo scaling. Make multiplicative models additive	Reduce multiplicative effects become additive	Difficulties with values with large relative standard deviation and zeros
Vast scaling	(-)	Focus on the metabolites that show small fluctuations	Aims for robustness, can use prior group knowledge	Not suited for large induced variation without group structure
Log pareto scaling	Log (-)	Reduce the relative importance of large values, but keep data structure partially intact	Stays closer to the original measurement than autoscaling	Sensitive to large fold changes

*van den Berg, R. A., Hoefsloot, H. C., Westerhuis, J. A., Smilde, A. K., & van der Werf, M. J. (2006). Centering, scaling, and transformations: improving the biological information content of metabolomics data. BMC genomics, 7(1), 142.*

User can check the normalization effect on the data for all methods by visualising the interactive boxplots tabs that are in **"Normalized Data"** panel. As more similar are the

# Statistical Analysis

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

The screenshot displays the POMA v1.0 web application interface. On the left is a dark blue sidebar with navigation options: Home, Input Data, Pre-processing, Statistics, Univariate analysis (highlighted), Multivariate analysis, Correlation analysis, Feature Selection, Random Forest, Rank Products, Automatic Statistical Analysis (highlighted with a red box), Help, Terms & Conditions, About Us, and Give us feedback. The main content area is divided into two panels. The left panel, titled 'Univariate methods:', contains settings for statistical tests: Limma, T-test (selected), ANOVA, Mann-Whitney U Test, and Kruskal Wallis Test. Below this are 'Variances are equal:' (TRUE, FALSE (selected)), 'Paired samples:' (TRUE, FALSE), and 'Volcano Plot Parameters:' with 'P.Value threshold' set to 0,05 and 'Fold change threshold' set to 1,5. The right panel, titled 'Results', shows a table of results with columns for Mean G1, Mean G2, FC (Ratio), Difference of Means, and P.V. The table lists several compounds like LinolenicAcid, Histidine, Deoxyuridine, PEP, MalonicAcid/3HBA, Glutamine, Methionine, and LinoleicAcid. Above the table are tabs for 'Results' and 'Volcano Plot', and buttons for 'Copy', 'Print', and 'Download' (highlighted with a red box). A search bar is also present.

	Mean G1	Mean G2	FC (Ratio)	Difference of Means	P.V
LinolenicAcid	-0.413	0.413	-1.000	-0.826	1.1006
Histidine	-0.315	0.315	-1.000	-0.630	2.2204
Deoxyuridine	-0.310	0.310	-1.000	-0.620	2.8004
PEP	-0.296	0.296	-1.000	-0.592	5.6504
MalonicAcid/3HBA	-0.293	0.293	-1.000	-0.586	6.5504
Glutamine	-0.261	0.261	-1.000	-0.522	2.4503
Methionine	-0.254	0.254	-1.000	-0.508	3.1803
LinoleicAcid	-0.250	0.250	-1.000	-0.500	3.8403
					4.42

# 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 could 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](mailto:polcaes@gmail.com)

 [pcastellanoescuder.github.io/](https://pcastellanoescuder.github.io/)

 [@polcastellano\\_](https://twitter.com/polcastellano_)

 [@pcastellanoescuder](https://github.com/pcastellanoescuder)

 [University of Barcelona](#)