

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

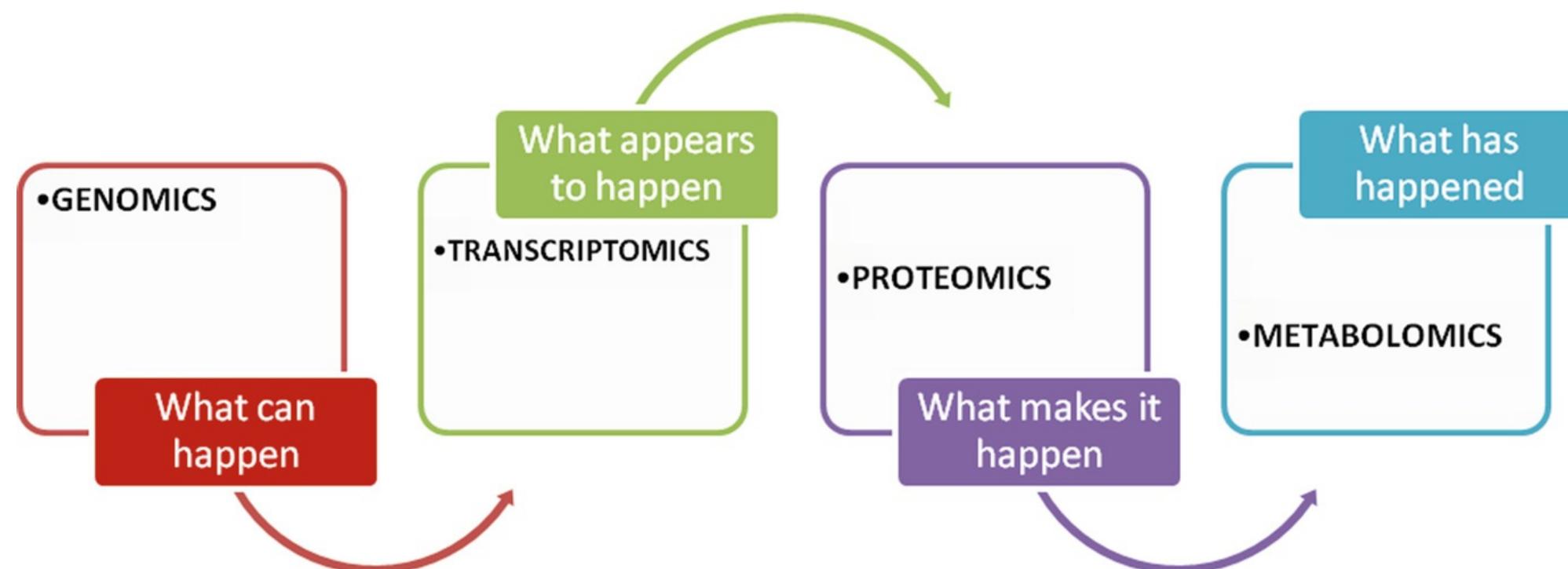


CONTEXT

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

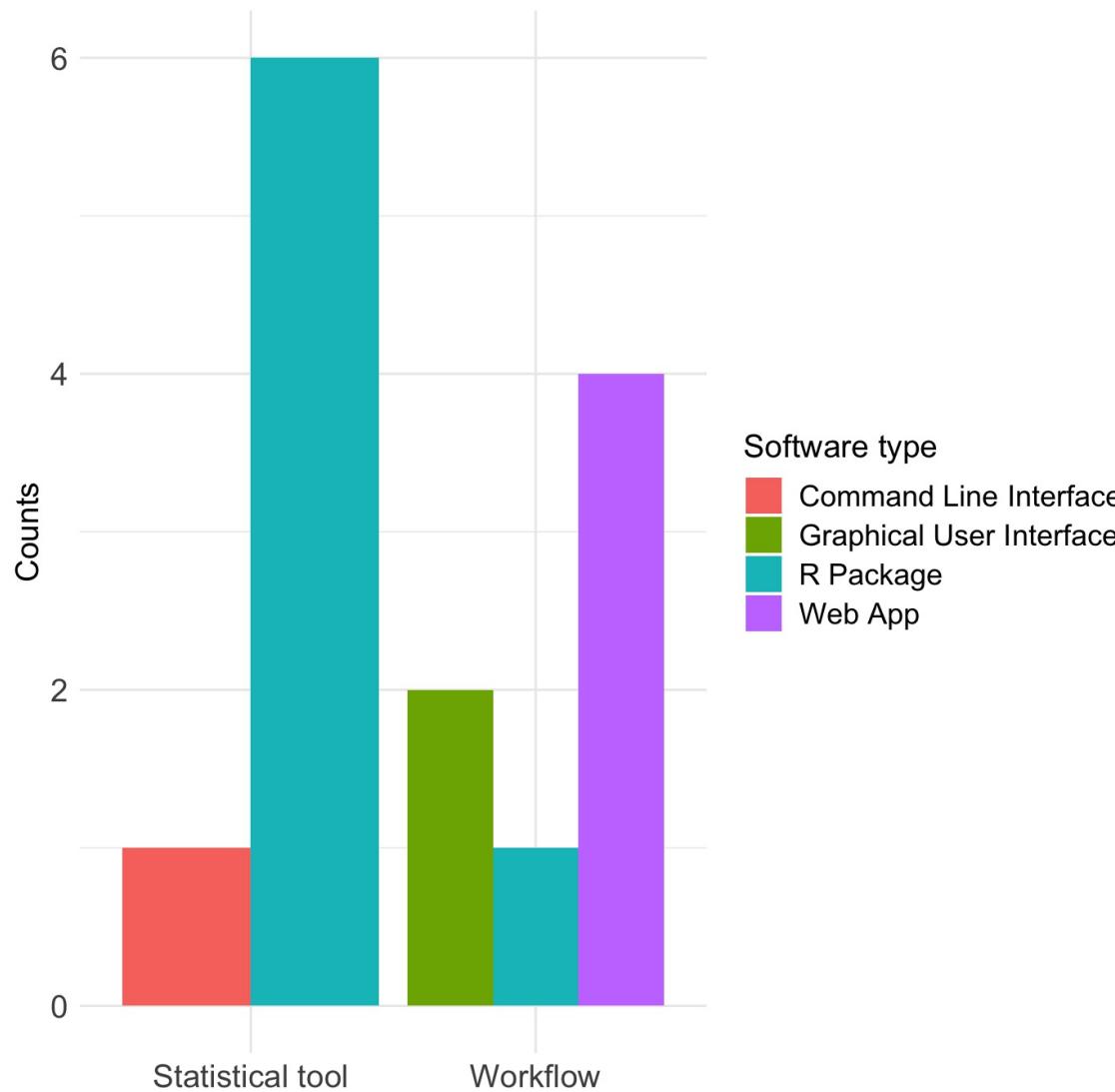
How is the data?

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

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)

Freely Available Existing Tools



Web Apps that allows users to perform a statistical analysis³

- Workflow4metabolomics
- Galaxy-M
- XCMS Online
- MetaboAnalyst

[3] Spicer, R., Salek, R. M., Moreno, P., Cañuelo, 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 **REPLICABLE** 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

 UNIVERSITAT DE
BARCELONA




Centro de Investigación Biomédica en Red
Fragilidad y Envejecimiento Saludable



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).

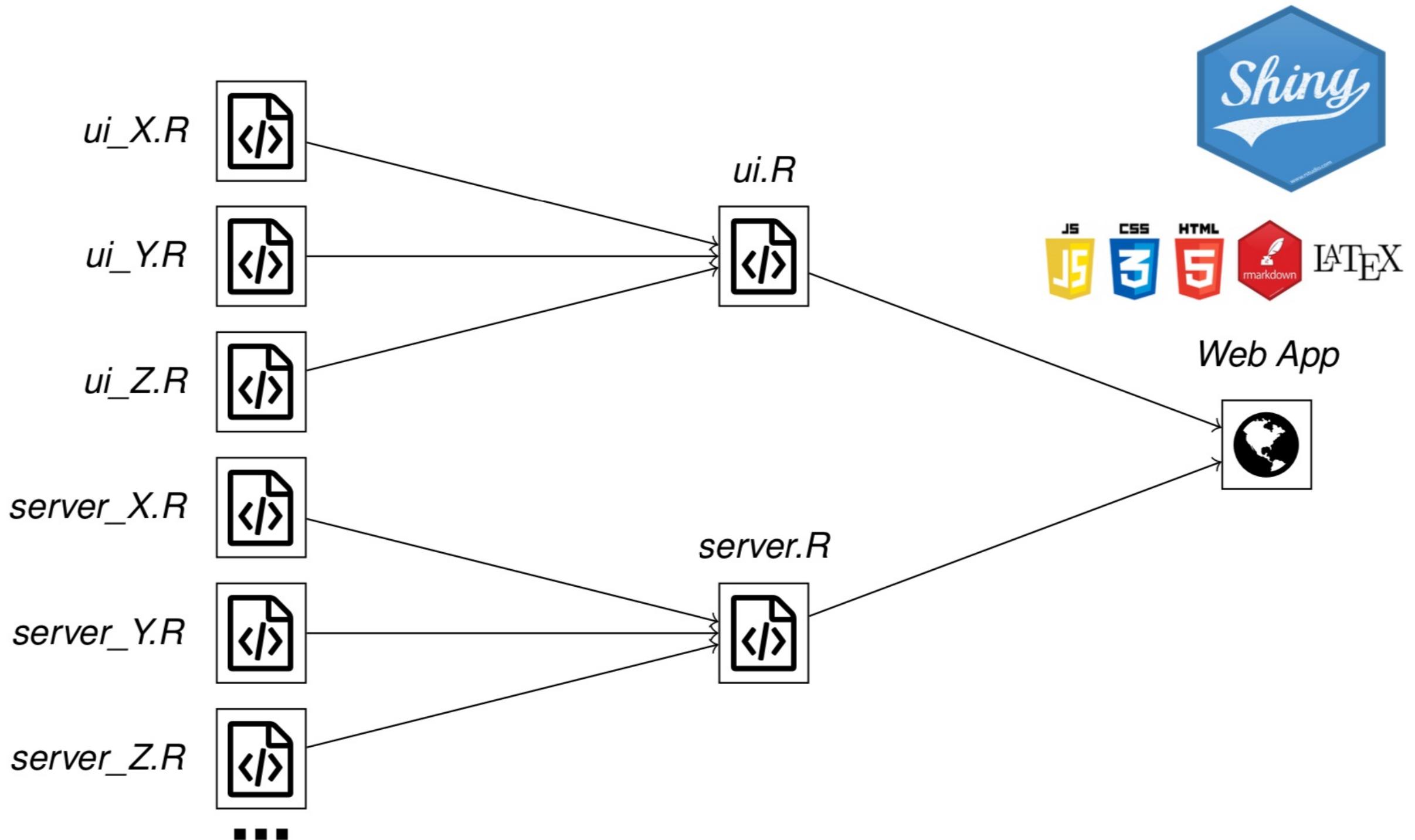
Metabolomic Data

- 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

The screenshot shows the POMA Input Data Panel. The left sidebar has a dark blue header with "POMA v1.0" and a teal footer. The main menu items are Home, Input Data (selected), Pre-processing, Statistics, Help, Terms & Conditions, About Us, and Give us feedback. The Input Data section is highlighted in orange. The sub-sections under Input Data are Samples (IDs), Groups, First Metabolite (set to Methyladenosine), and Last Metabolite (set to Xanthurenone). A red "Submit" button is at the bottom. A note below it says: "After click the button above, go to the Pre-processing step". On the right, there's a green header "Uploaded Data" with a "Exploratory report" button. Below is a table showing 10 entries of uploaded data. The table has columns: ID, Groups, Methyladenosine, Methylhistamine, Aminoacidate, Deoxyuridine, and Nitrotyrosine. The table shows data for metabolites like C, 157, 200, etc. At the bottom of the table are navigation links: Previous, 1, 2, 3, 4, 5, ..., 14, Next.

ID	Groups	Methyladenosine	Methylhistamine	Aminoacidate	Deoxyuridine	Nitrotyrosine
157	C	363294	17961	211814	13208	581
200	C	258237	42811	129058	12801	521
133	C	414501	27449	419827	15744	838
250	C	176266	31305	74720	12989	448
109	C	390954	34627	141257	13116	929
77	C	335439	26145	139377	8096	101:
177	C	485946	48012	182545	13856	945
132	C	412251	44478	235936	15693	89:
257	C	475436	27005	192159	13499	124:
173	C	572292	24994	315960	11378	778

Input Data Panel

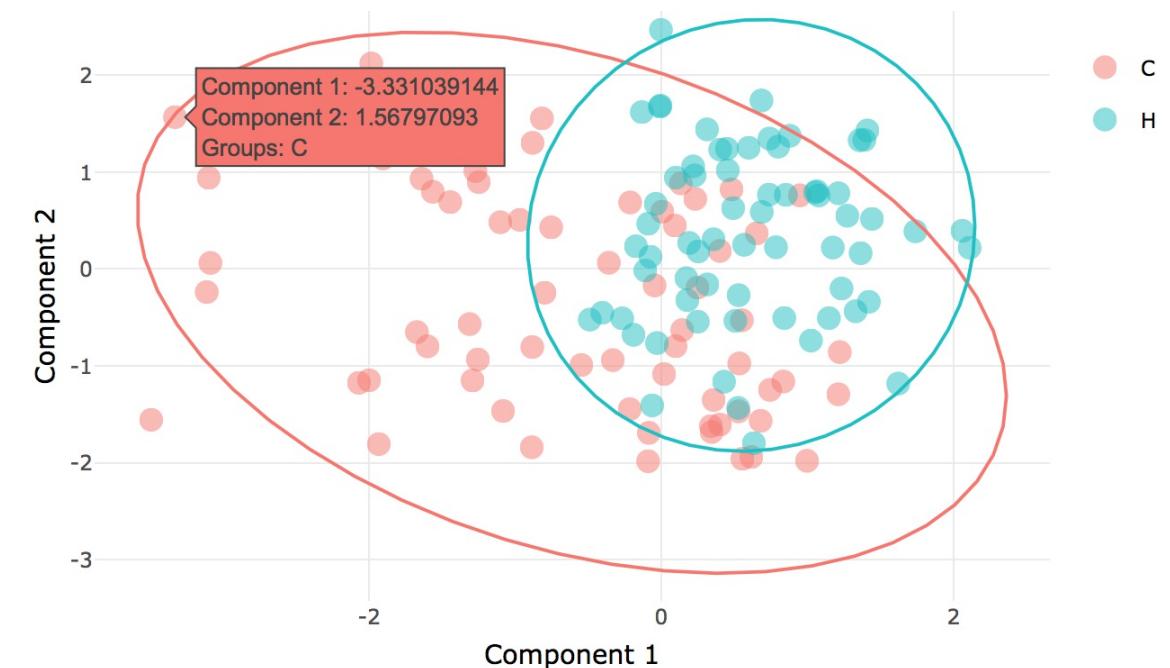
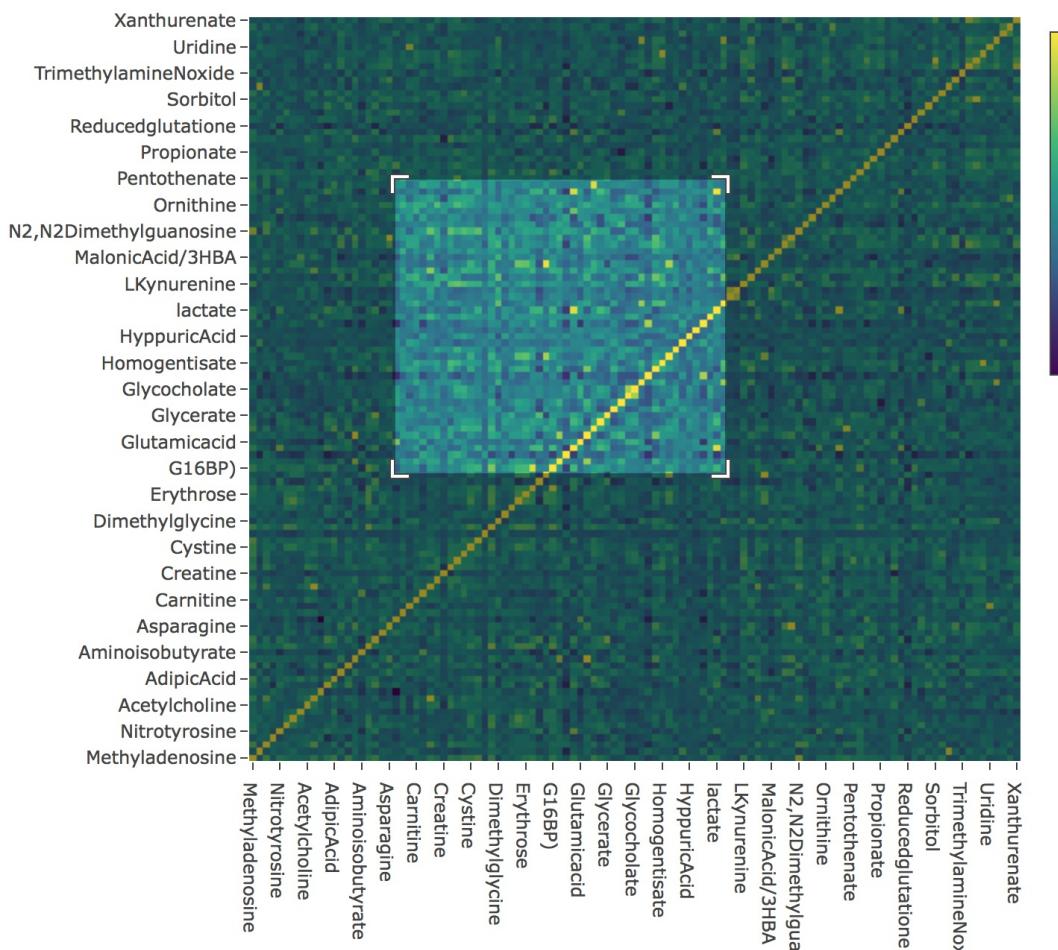
We have used the `shinydashboard` package for the main structure and the `dashboardthemes` package for customization

The screenshot shows the POMA v1.0 Input Data Panel. The left sidebar has links for Home, Input Data (which is selected), Pre-processing, Statistics, Help, Terms & Conditions, About Us, and Give us feedback. The main area has a green header bar with a menu icon. Below it, a red box highlights the "Exploratory report" button. A red box also highlights the "Do you want to use our example data?" section, which contains two radio buttons: "Yes" (selected) and "No, upload my own data". Another red box highlights the "Samples (IDs)" dropdown set to "ID". A red box highlights the "Groups" dropdown set to "Groups". A red box highlights the "First Metabolite" dropdown set to "Methyladenosine". A red box highlights the "Last Metabolite" dropdown set to "Xanthurenone". A red box highlights the "Submit" button. A red box highlights the "Covariates file" input field. The central part of the screen shows a table titled "Uploaded Data" with columns: ID, Groups, Methyladenosine, Methylhistamine, Amino adipate, Deoxyuridine, and Nitrotyrosine. The table shows 10 entries out of 132. The bottom of the screen shows "Prepared Data" and a red box highlights the "Covariates file" input field.

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

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 shows the POMA v1.0 web application interface. The left sidebar contains navigation links: Home, Input Data, Pre-processing, Impute Values, Normalization (which is highlighted in orange), Statistics, Help, Terms & Conditions, About Us, and Give us feedback. The main content area has a header "Normalization methods:" and a list of normalization options: None, Autoscaling, Level scaling, Log scaling (selected), Log transformation, Vast scaling, and Log pareto scaling. Below this is a green "Normalize" button and a blue help icon with a question mark, which is circled in red. To the right, there are two tabs: "Not Normalized Data" (blue background) and "Normalized Data" (green background). Under "Normalized Data", there are three sub-tabs: Data, Raw Data Boxplot, and Normalized Boxplot. Below these are buttons for Copy, Print, and Download, and a search bar labeled "Search:". A table follows, with columns ID, Group, Methyladenosine, Methylhistamine, Amino adipate, and Deoxyuridine. The table rows are:

ID	Group	Methyladenosine	Methylhistamine	Amino adipate	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

The screenshot shows the POMA app interface. The top navigation bar includes 'POMA v1.0', a user icon, and a search bar. The left sidebar has links for Home, Input Data, Pre-processing, Impute Values, Normalization (which is selected and highlighted in orange), Statistics, Help, Terms & Conditions, About Us, and Give us feedback. The main content area shows a 'Normalization helper' panel with the following text:

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	UnitGoal	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 transformationO	Reduce heteroscedasticity, multiplicative effects standard deviation and zeros	Difficulties with values with large relative
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

POMA v1.0

Home
Input Data
Pre-processing
Statistics
» Univariate analysis
» Multivariate analysis
» Correlation analysis
» Feature Selection
» Random Forest
» Rank Products
★ Automatic Statistical Analysis

Help
Terms & Conditions
About Us
Give us feedback

Univariate methods:
 Limma
 T-test
 ANOVA
 Mann-Whitney U Test
 Kruskal Wallis Test

Variances are equal:
 TRUE
 FALSE (Welch's T-test)

Paired samples:
 TRUE
 FALSE

Volcano Plot Parameters:
P.Value threshold: 0,05
Fold change threshold: 1,5
xlim range: 1 2 10

Results Volcano Plot

Copy Print Download

Search:

	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 HETROSCEDASTICITY
 - 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

🔗 pcastellanoescuder.github.io/

🐦 @polcastellano_

🗣 @pcastellanoescuder

📍 University of Barcelona