

# POMA Shiny Framework

## Statistical analysis tool for targeted metabolomic data

**Pol Castellano-Escuder<sup>1,2,3</sup>**

<sup>1</sup> Biomarkers and Nutritional & Food Metabolomics Research Group,  
Department of Nutrition, Food Science and Gastronomy. University of Barcelona.

<sup>2</sup> Statistics and Bioinformatics Research Group,  
Department of Genetics, Microbiology and Statistics. University of Barcelona.

<sup>3</sup> CIBERFES, Carlos III Health Institute.

25 Jan 2019

# Overview

What's metabolomics ?

# Overview

## 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."*

# Overview

## 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."*

## Common pipeline analysis for target metabolomics data

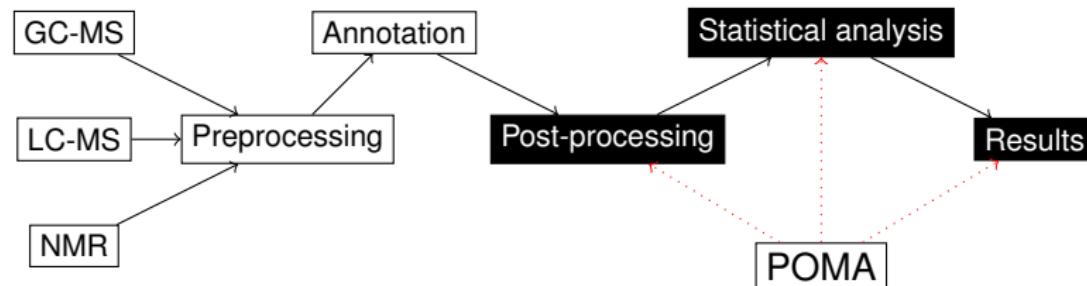
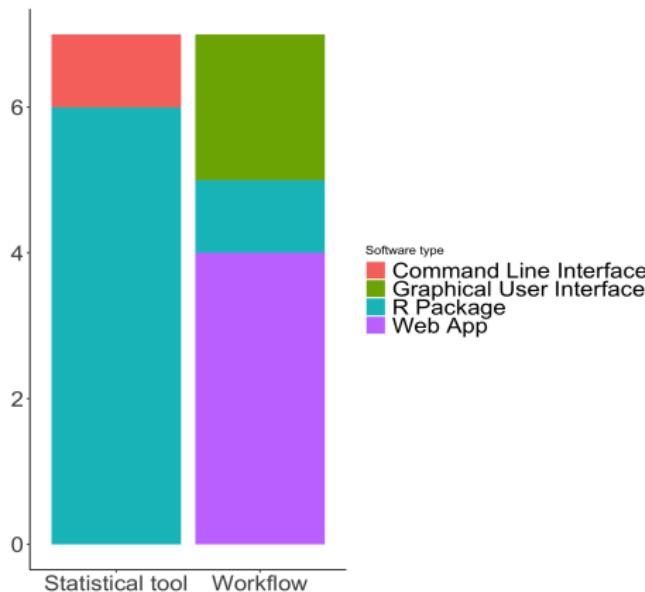


Diagram of pipeline analysis for target metabolomic data

# Existing Tools



Web Apps that allows users to perform a statistical analysis

- Workflow4metabolomics
- Galaxy-M
- XCMS Online
- MetaboAnalyst 3.0

*Rachel S., et. al. Metabolomics, 2017.*

# Motivation & Aims

## Motivation

# 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.

# 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.

# 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

# 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.

# 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).

# 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)

# 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.

# 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.

## Concept

- In this case, the main aim is **COMPLETE** and give other option to users, NOT to **COMPETE** with the existing tools.

# Methods

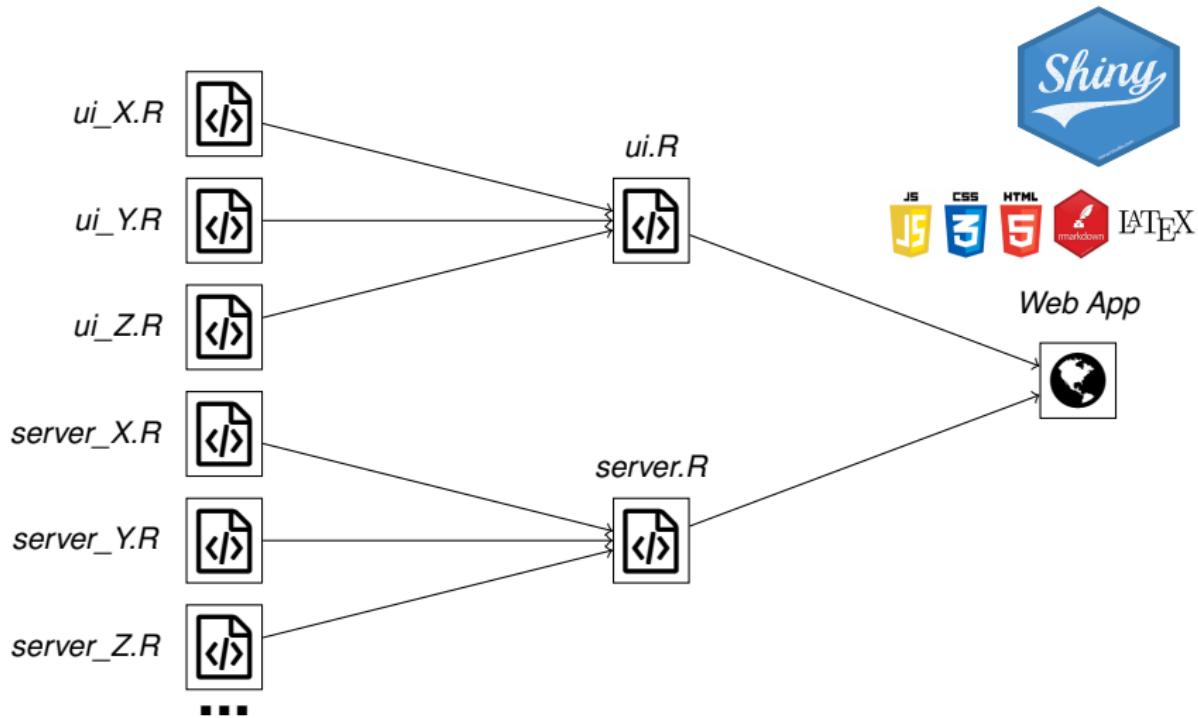


Diagram of POMA files architecture

# POMA Web 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**

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 :

[polcastellano.shinyapps.io/POMA/](http://polcastellano.shinyapps.io/POMA/)

# Input Data

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

Xanthurene

Submit

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

Uploaded Data

Show 10 entries

ID	Groups	Methyladenosine	Methylhistamine	Aminoadipate	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

POMA offers a colorectal carcinoma (CRC) metabolomic dataset as example to test the app without uploading your own data.

# Input Data

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

First Metabolite

Methyladenosine

Last Metabolite

Xanthureneate

Submit

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

Uploaded Data

Show: 10 entries

Search:

ID	Groups	Methyladenosine	Methyhistamine	Aminoadipate	Deoxyuridine	Nitrotyros
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

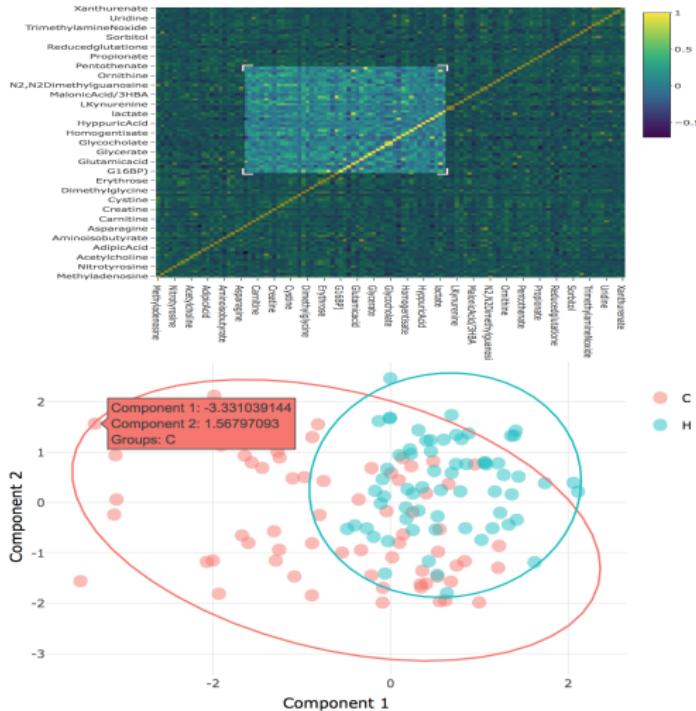
Previous 1 2 3 4 5 ... 14 Next

Prepared Data

Covariates file

POMA offers a colorectal carcinoma (CRC) metabolomic dataset as example to test the app without uploading your own data.

# Visualization



All plots in the app are designed using **plotly** library. This library make plots interactive.

Users can 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.

# Helpers

The screenshot shows the POMA v1.0 interface. On the left, a sidebar menu includes Home, Input Data, Pre-processing (with 'Normalization' selected), Impute Values, Statistics, Help, Terms & Conditions, About Us, and Give us feedback. The main area has a title 'Normalization methods:' with several options: None, Autoscaling, Level scaling, Log scaling (which is selected), Log transformation, Vast scaling, and Log pareto scaling. Below this are two buttons: 'Normalize' and a question mark icon. To the right, there are two tabs: 'Not Normalized Data' and 'Normalized Data'. Under 'Normalized Data', there are three sub-tabs: Data, Raw Data Boxplot, and Normalized Boxplot. Below these tabs is a search bar labeled 'Search:'. The main content area displays a table of data with columns: ID, Group, Methyladenosine, Methylhistamine, Aminoacidate, and Deoxyuridine. The data rows are as follows:

ID	Group	Methyladenosine	Methylhistamine	Aminoacidate	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

Every panel include an individual **Help Panel**.

# Helpers

**POMA v1.0**

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

**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	UnitGoal	Advantages	Disadvantages
Autoscaling	(-) Compare metabolites based on correlations	All metabolites become equally important	Inflation of the measurement errors
Level scaling	(-) Focus on relative response	Suited for identification of e.g. biomarkers	Inflation of the measurement errors
Log scaling	Log (-)	Correct for Log heteroscedasticity, pseudo transformation	Difficulties with values with large relative standard deviation and zeros
Log transformation	Log	Scaling. Make multiplicative models 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 (-) but keep data structure partially intact	Reduce the relative importance of large values, 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

Search:	Aminoadipate	Deoxyuridine
39	4.188	2.438
33	2.442	0.533
43	0.196	-0.003
72	0.694	0.782
01	-0.03	-0.032
53	2.832	0.5
91	1.766	-1.409
13	1.453	0.847
28	1.856	1.8
33	0.6	-0.604
06	-0.171	-0.432
99	-0.525	0.524

## Normalization tab helper.

# Results

The screenshot shows the POMA v1.0 Shiny application interface. On the left, a sidebar menu includes Home, Input Data, Pre-processing, Statistics (with Univariate analysis selected), Multivariate analysis, Correlation analysis, Feature Selection, Random Forest, Rank Products, Automatic Statistical Analysis, Help, Terms & Conditions, About Us, and Give us feedback. The main panel displays a list of biomarkers with their statistical parameters. At the top right, there are buttons for Results, Volcano Plot, Copy, Print, Download, and a search bar. The table below lists the following data:

	Mean G1	Mean G2	FC (Ratio)	Difference of Means	P.V
LinolenicAcid	-0.413	0.413	-1.000	-0.826	1.10 06
Histidine	-0.315	0.315	-1.000	-0.630	2.24 04
Deoxyuridine	-0.310	0.310	-1.000	-0.620	2.80 04
PEP	-0.296	0.296	-1.000	-0.592	5.65 04
MalonicAcid/3HBA	-0.293	0.293	-1.000	-0.586	6.55 04
Glutamine	-0.261	0.261	-1.000	-0.522	2.45 03
Methionine	-0.254	0.254	-1.000	-0.508	3.16 03
LinoleicAcid	-0.250	0.250	-1.000	-0.500	3.84 03
					4.43

*The understandable results can be filtered, columns re-ordered and downloaded as .PDF, .CSV or Excel.*

# Results

The screenshot shows the POMA v1.0 application interface. On the left, there's a sidebar with navigation links: Home, Input Data, Pre-processing, Statistics (with Univariate analysis selected), Multivariate analysis, Correlation analysis, Feature Selection, Random Forest, Rank Products, and Automatic Statistical Analysis (which is highlighted with a red box). Below these are Help, Terms & Conditions, About Us, and Give us feedback.

The main content area has a title "Univariate methods:" followed by a list of options: Limma, T-test (which is selected and highlighted with a blue circle), ANOVA, Mann-Whitney U Test, and Kruskal Wallis Test.

Below that is a section "Variances are equal:" with options TRUE (selected) and FALSE (Welch's T-test).

Then there's a section "Paired samples:" with options TRUE (selected) and FALSE.

Underneath these sections is a "Volcano Plot Parameters:" group with three input fields: P.Value threshold (0,05), Fold change threshold (1,5), and xlim range (1 2 10).

To the right of this is a table titled "Results" with a "Volcano Plot" tab selected. The table has columns: Mean G1, Mean G2, FC (Ratio), Difference of Means, and P.V. The table lists several metabolites with their respective values:

	Mean G1	Mean G2	FC (Ratio)	Difference of Means	P.V
LinolenicAcid	-0.413	0.413	-1.000	-0.826	1.10 06
Histidine	-0.315	0.315	-1.000	-0.630	2.22 04
Deoxyuridine	-0.310	0.310	-1.000	-0.620	2.86 04
PEP	-0.296	0.296	-1.000	-0.592	5.65 04
MalonicAcid/3HBA	-0.293	0.293	-1.000	-0.586	6.55 04
Glutamine	-0.261	0.261	-1.000	-0.522	2.45 03
Methionine	-0.254	0.254	-1.000	-0.508	3.18 03
LinoleicAcid	-0.250	0.250	-1.000	-0.500	3.84 03
					4.45

*The automatic statistical analysis generate an univariate specific-analysis according to variable conditions.*

# Conclusions

- 1) We have developed a **FAST, FRIENDLY** and **FREE** software that is called **POMA**.

# Conclusions

- 1) We have developed a **FAST, FRIENDLY** and **FREE** software that is called **POMA**.
- 2) POMA is based in **R language** and uses a **Shiny** system to run.

# Conclusions

- 1) We have developed a **FAST, FRIENDLY** and **FREE** software that is called **POMA**.
- 2) POMA is based in **R language** and uses a **Shiny** system to run.
- 3) POMA provides an accurate documentation ("HELP") at each step of analysis that could improve the results and facilitate the interpretation of it.

# Conclusions

- 1) We have developed a **FAST, FRIENDLY** and **FREE** software that is called **POMA**.
- 2) POMA is based in **R language** and uses a **Shiny** system to run.
- 3) POMA provides an accurate documentation ("HELP") at each step of analysis that could improve the results and facilitate the interpretation of it.
- 4) POMA can generate two types of **automatic reports** : **Exploratory report** and **Statistical report**.

# Conclusions

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

## Group

Cristina Andrés  
Alexandre Sánchez

Ferran Reverter  
Esteban Vegas  
Antonio Mlñarro  
Francesc Carmona  
Marta Cubedo

Raúl González  
Gregorio Peron  
Nicola Kielland  
Nicole Hidalgo  
Magalí Palau  
Marjorie Reyes  
Núria Estanyol  
María Cristina Cadena

---

## Contact

- ✉ polcaes@gmail.com
- ✉ polcastellano\_
- ✉ pcastellanoescuder
- ✉ <https://pcastellanoescuder.github.io>

## Institutions



Centro de Investigación Biomédica en Red  
Fragilidad y Envejecimiento Saludable

