asterics

A tool for the exploration and integration of omics data

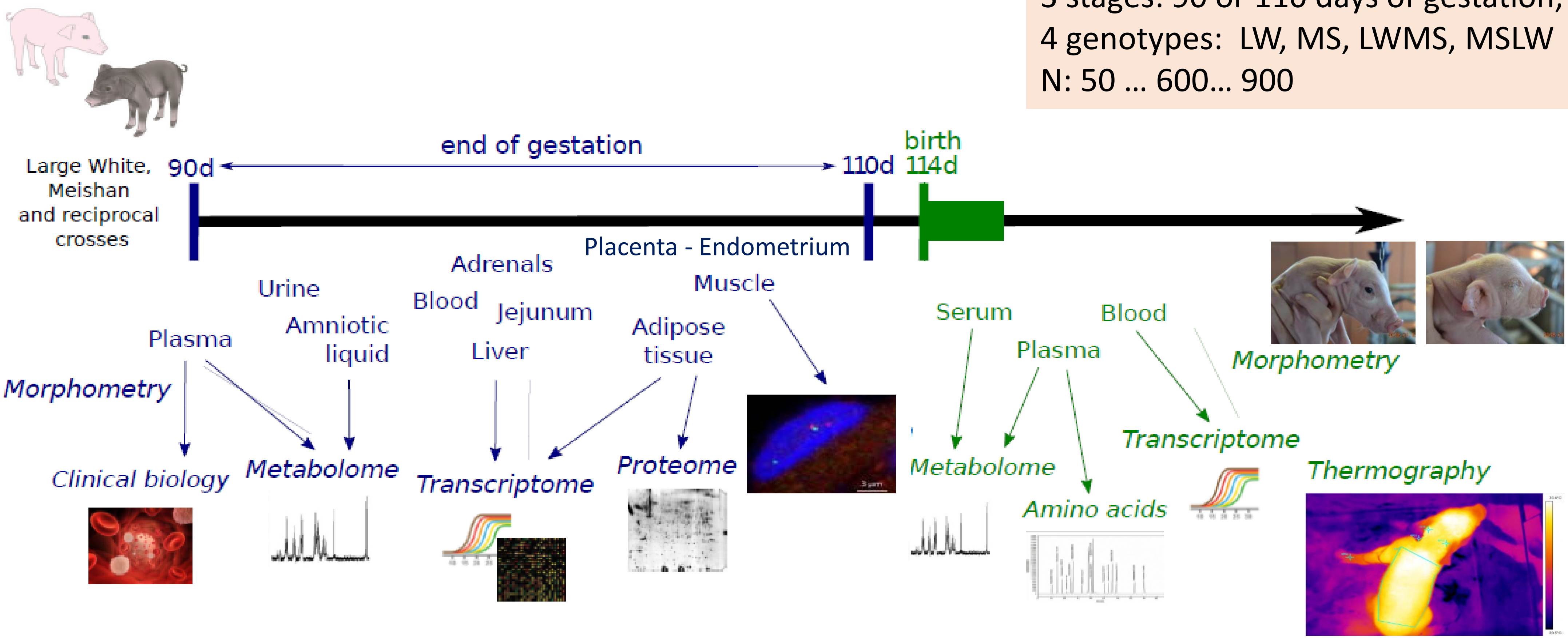
ASTERICS is an online tool designed to help you to perform your statistical and integrative analyses in an interactive and easy-to-use way.

Quelques exemples avec des données multi-omiques et multi-tissus en néonatalité porcine pour illustrer les différentes fonctionnalités d'Asterics.

Élise Maigné, Céline Noirot, Jérôme Mariette, Yaa Adu Kesewaah, Sébastien Déjean, Camille Guilmineau, Julien Henry, Arielle Krebs, **Laurence Liaubet**, Fanny Mathevet, Hyphen-Stat, Christine Gaspin, Nathalie Vialaneix

Context: neonatal maturity and birth survival

3 stages: 90 or 110 days of gestation, birth
4 genotypes: LW, MS, LWMS, MSLW
N: 50 … 600… 900



Large White, Meishan and reciprocal crosses

90d

end of gestation

110d

birth 114d

Morphometry

Plasma

Urine

Amniotic liquid

Blood

Adrenals

Liver

Jejunum

Placenta - Endometrium

Adipose tissue

Muscle

Clinical biology

Metabolome

Transcriptome

Proteome

Serum

Plasma

Blood

Metabolome

Amino acids

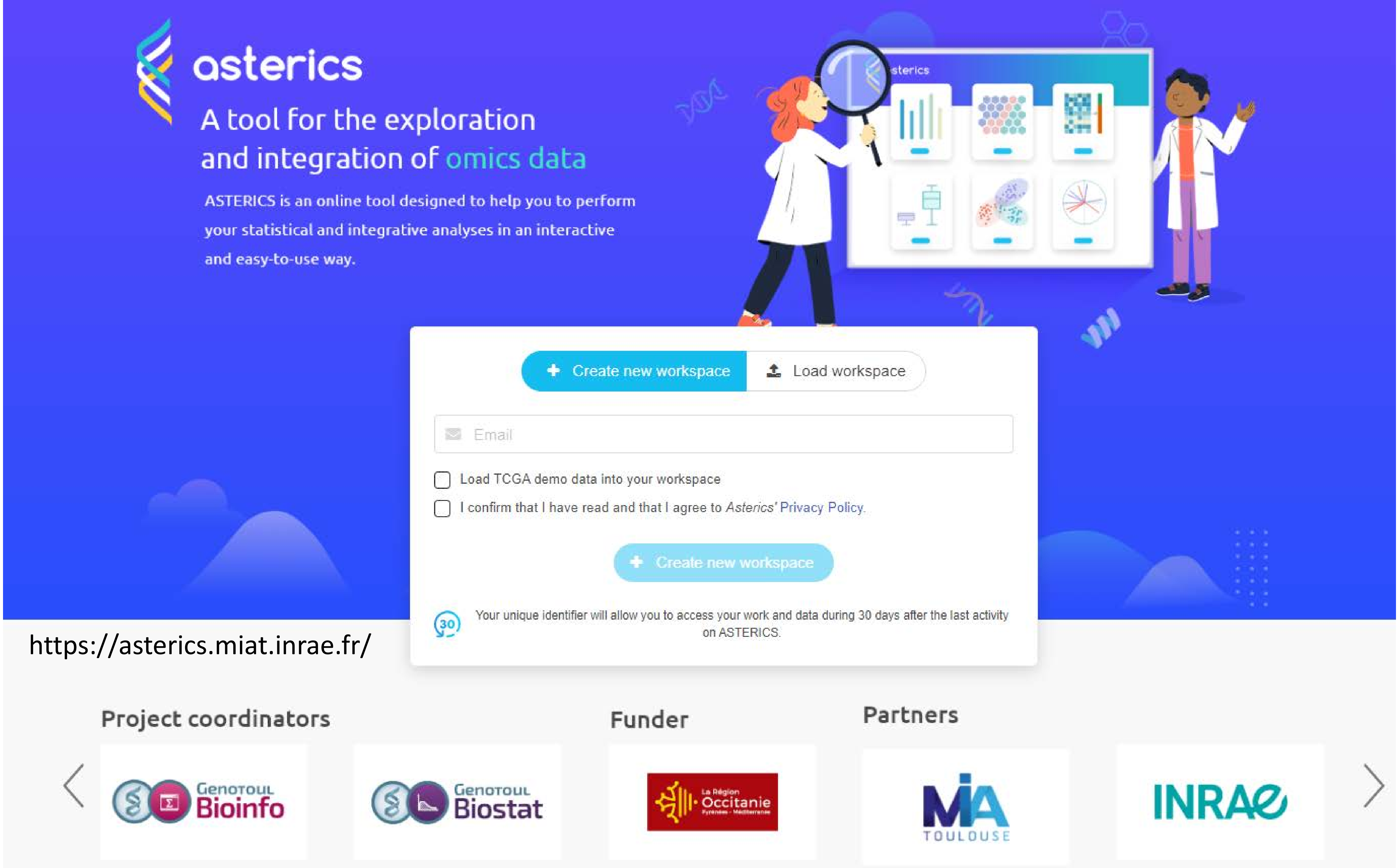
Transcriptome

Morphometry

Thermography

Voillet et al, 2014, 2018; Yao et al.2017 ; Gondret et al, 2018;
Marti-Marimon et al, 2018; Lefort et al, 2020,2021 ; Schmitt et al, 2021

Avec la contribution financière du compte d'affectation spéciale développement agricole et rural CASDAR
MINISTÈRE DE L'AGRICULTURE ET DE L'ALIMENTATION
ALLIANCE R&D Innovation Durable en Génétique Porcine
INRA SCIENCE & IMPACT Génétique Animale
NeoTec-Vision
ifip institut du porc



asterics

A tool for the exploration and integration of omics data

ASTERICS is an online tool designed to help you to perform your statistical and integrative analyses in an interactive and easy-to-use way.

☐ Load TCGA demo data into your workspace

☐ I confirm that I have read and that I agree to *Asterics'* Privacy Policy.

Your unique identifier will allow you to access your work and data during 30 days after the last activity on ASTERICS.

https://asterics.miat.inrae.fr/

Project coordinators

Funder

Partners

Dear user,

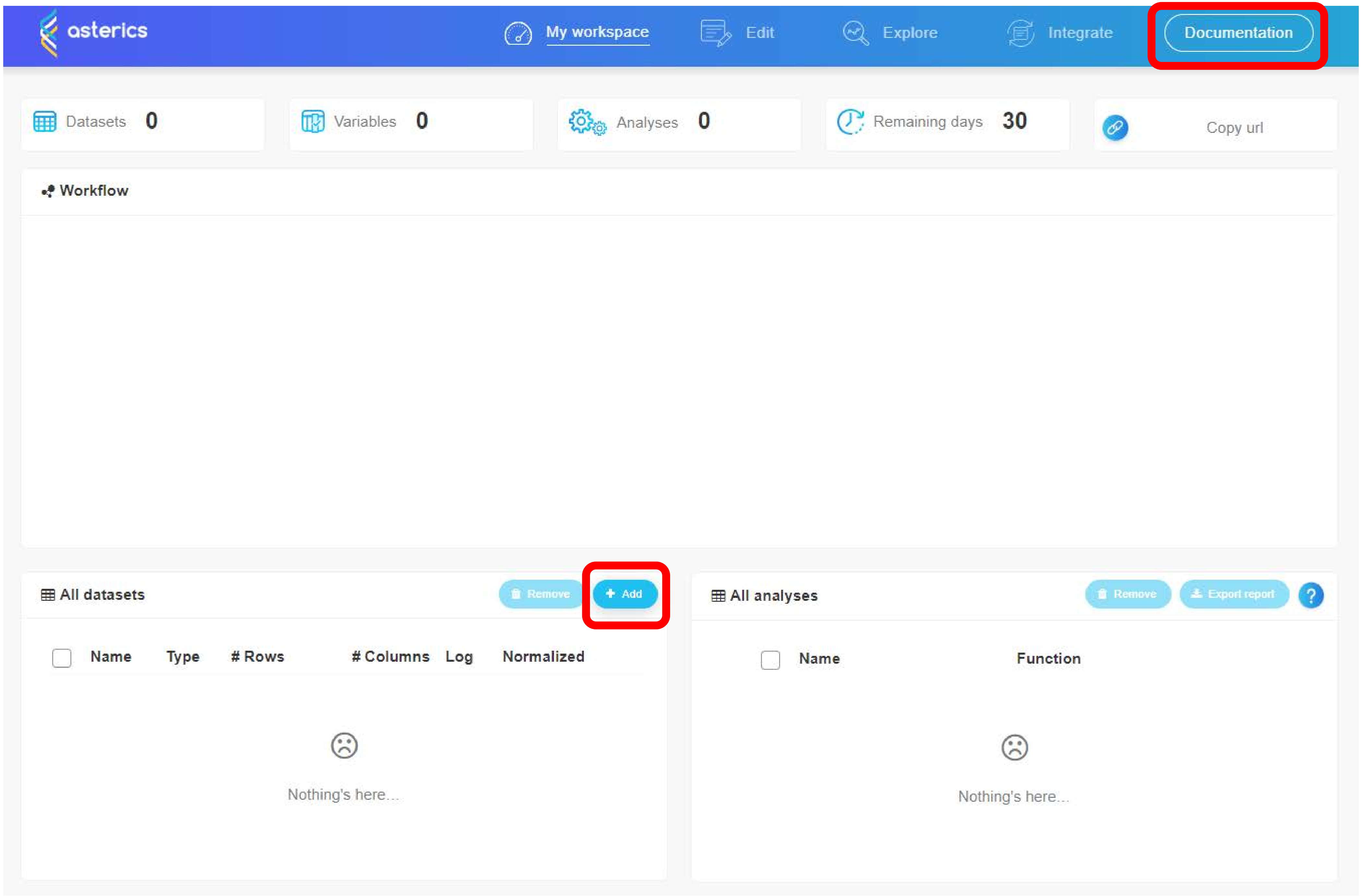
Your workspace has been successfully created at
https://asterics.miat.inrae.fr/workspace/ba20fb6d-4196-45c0-999f-7ff6d522b109.

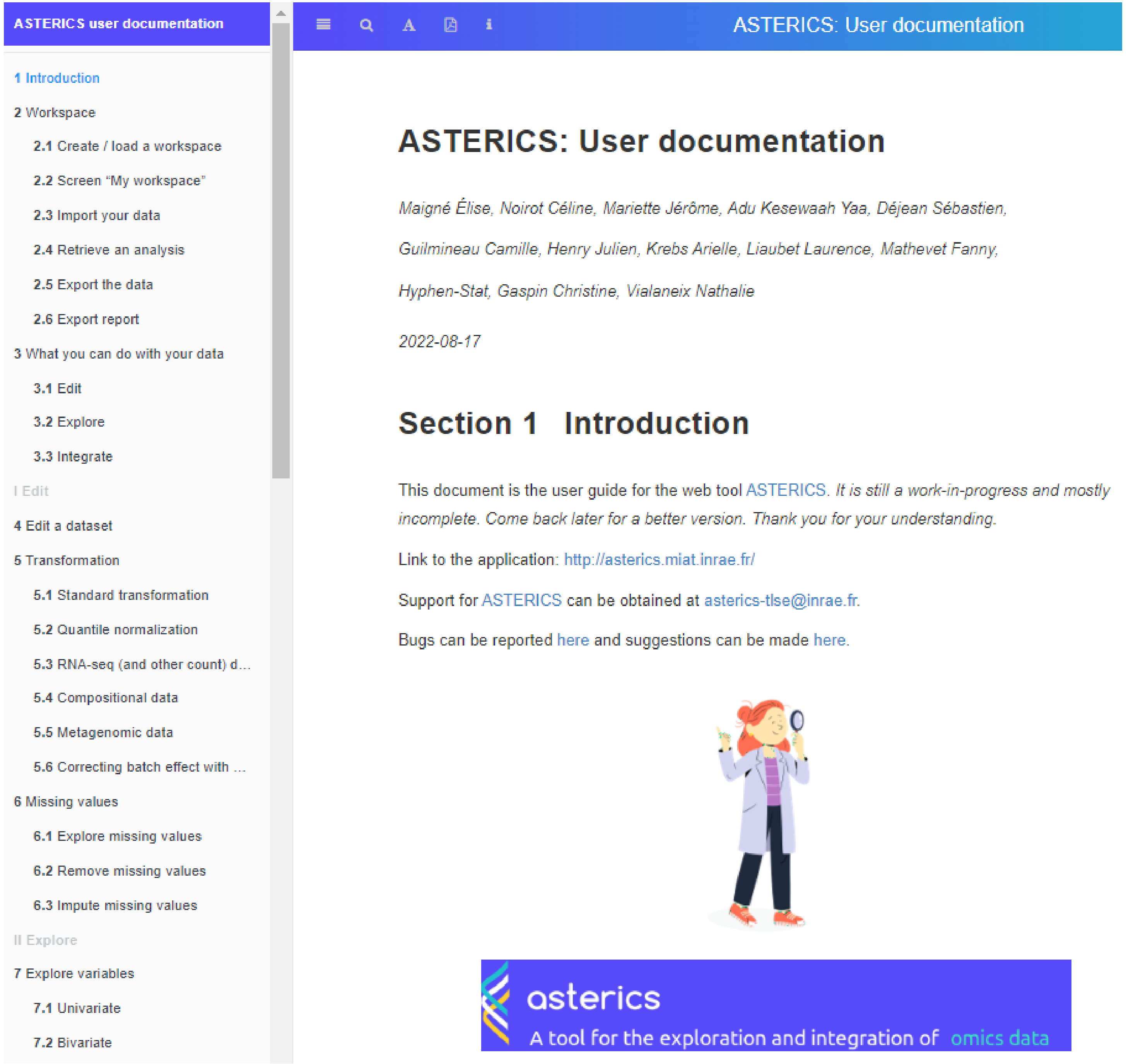
Thank you for using ASTERICS,
The ASTERICS Team

Dear user,

your workspace https://asterics.miat.inrae.fr/78f3057f-ed70-42c8-ae5c-8d8b7be83484 will be deleted in 3 days.

thanks for using Asterics,
The Asterics Team,


asterics
My workspace
Edit
Explore
Integrate
Documentation
Datasets 0
Variables 0
Analyses 0
Remaining days 30
Copy url
Workflow
All datasets
Remove
+ Add
☐ Name Type # Rows # Columns Log Normalized
Nothing's here...
All analyses
Remove
Export report
?
☐ Name Function
Nothing's here...



ASTERICS: User documentation

Maigné Élise, Noirot Céline, Mariette Jérôme, Adu Kesewaah Yaa, Déjean Sébastien,

Guilmineau Camille, Henry Julien, Krebs Arielle, Liaubet Laurence, Mathevet Fanny,

Hyphen-Stat, Gaspin Christine, Vialaneix Nathalie

2022-08-17

Section 1 Introduction

This document is the user guide for the web tool ASTERICS. *It is still a work-in-progress and mostly incomplete. Come back later for a better version. Thank you for your understanding.*

Link to the application: http://asterics.miat.inrae.fr/

Support for ASTERICS can be obtained at asterics-tlse@inrae.fr.

Bugs can be reported here and suggestions can be made here.

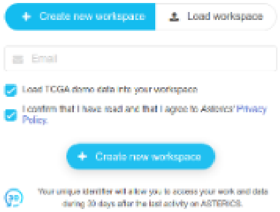
Section 16 Case studies

16.1 Breast cancer

This case study focuses on breast cancer, trying to link omics data to cancer subtypes. There are four breast cancer subtypes that somewhat represent sub-diseases, with different biological mechanisms. Their identification is particularly important when designing / selecting treatments. For this case study, we have mRNA and miRNA data on 988 individuals, as well as the associated subtypes. mRNA data is already normalized and is loaded in ASTERICS by default, whereas miRNA comes in the form of raw counts.
The data is available online (Vialaneix 2021).

16.1.1 Setup

Let us open an empty session and load the TCGA demo data.

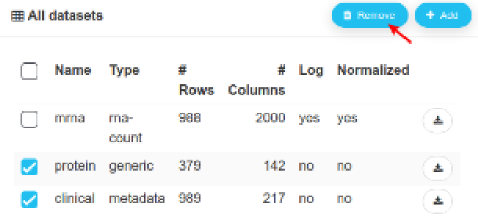


☑ Load TCGA demo data into your workspace

☑ I confirm that I have read and that I agree to Asterics' Privacy Policy.

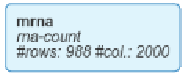
Your unique identifier will allow you to access your work and data during 30 days after the last activity on ASTERICS.

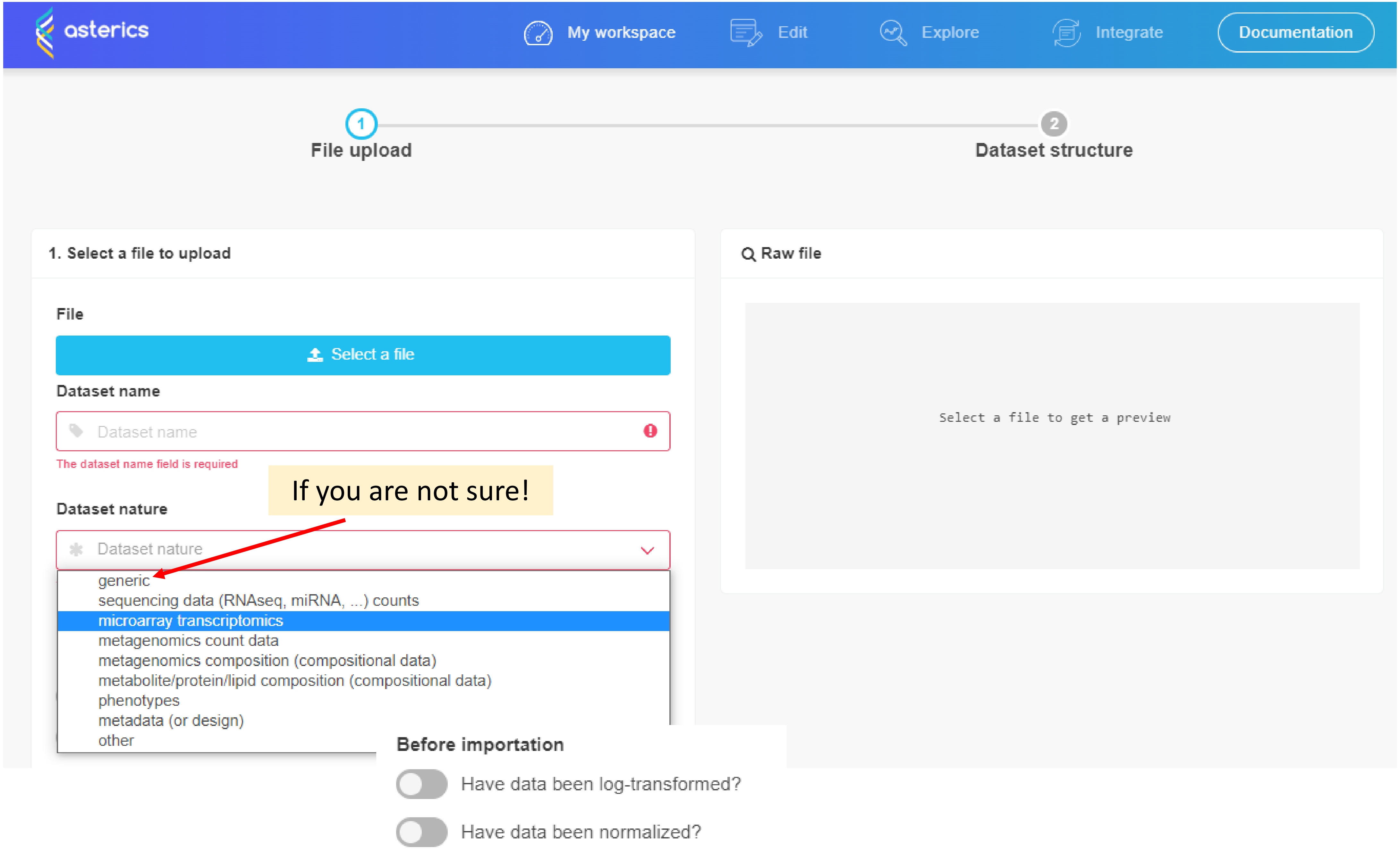
Since we will only use the **mrna** dataset, let us delete the other two.



☐	Name	Type	# Rows	# Columns	Log	Normalized
☐	mrna	rna-count	988	2000	yes	yes
☑	protein	generic	379	142	no	no
☑	clinical	metadata	989	217	no	no

Here is **mrna** in the workspace:





1. Select a file to upload

File

Dataset name

Dataset name

The dataset name field is required

If you are not sure!

Dataset nature

Dataset nature

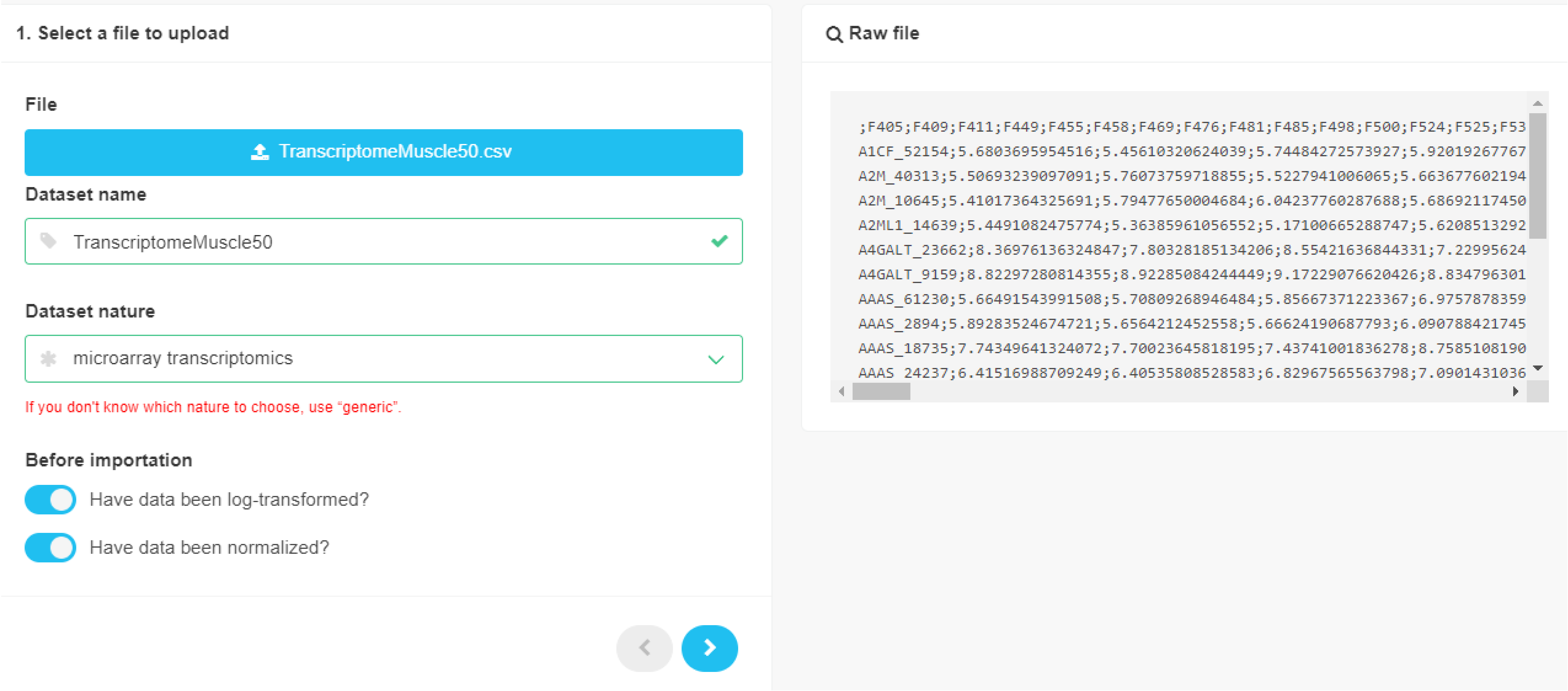
- generic
- sequencing data (RNAseq, miRNA, ...) counts
- microarray transcriptomics
- metagenomics count data
- metagenomics composition (compositional data)
- metabolite/protein/lipid composition (compositional data)
- phenotypes
- metadata (or design)
- other

Select a file to get a preview

Before importation

Have data been log-transformed?

Have data been normalized?



1. Select a file to upload

File

TranscriptomeMuscle50.csv

Dataset name

TranscriptomeMuscle50

Dataset nature

microarray transcriptomics

If you don't know which nature to choose, use "generic".

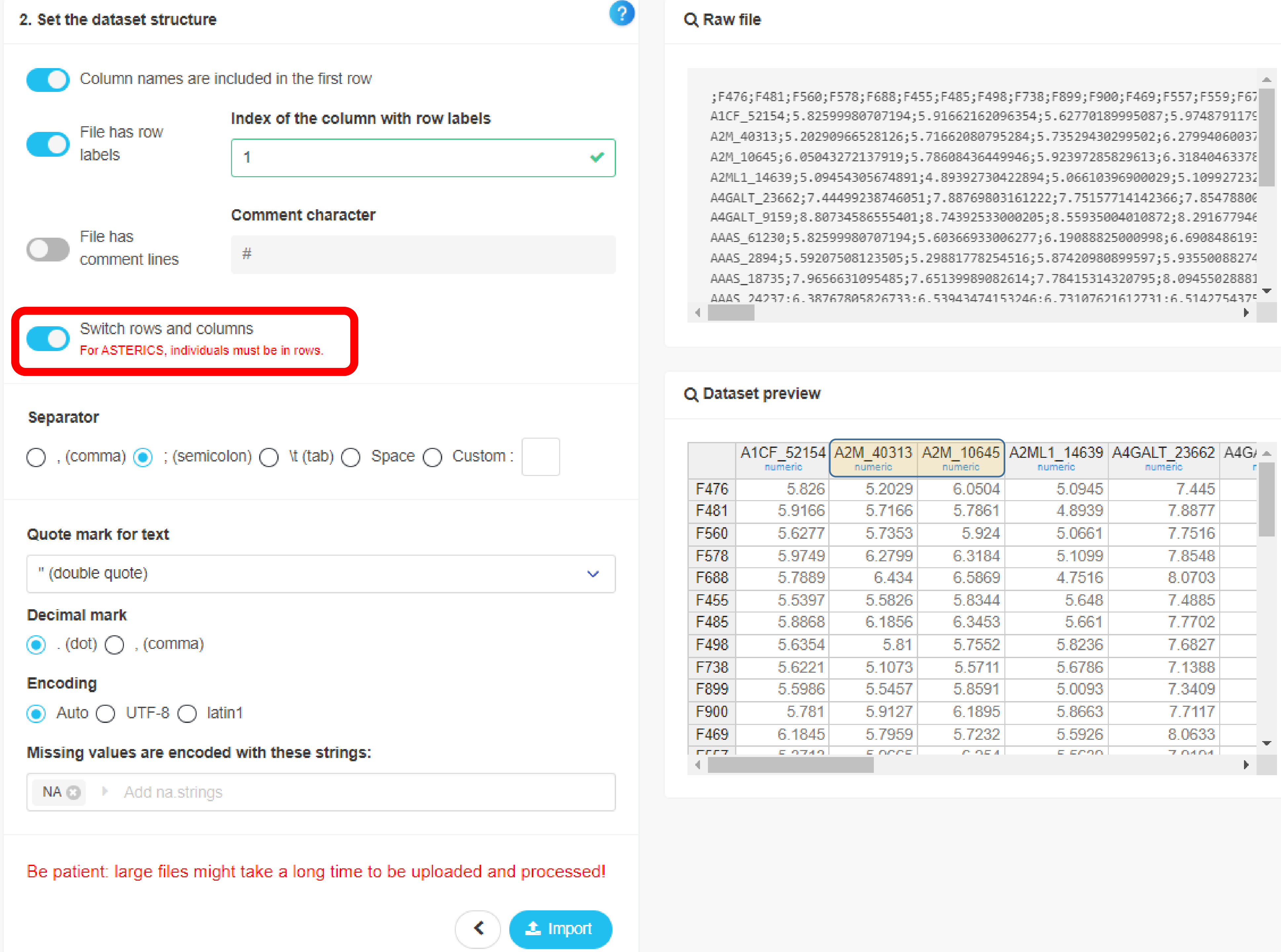
Before importation

Have data been log-transformed?

Have data been normalized?

Raw file

```
;F405;F409;F411;F449;F455;F458;F469;F476;F481;F485;F498;F500;F524;F525;F53
A1CF_52154;5.6803695954516;5.45610320624039;5.7448427257392 7;5.92019267767
A2M_40313;5.50693239097091;5.76073759718855;5.5227941006065;5.663677602194
A2M_10645;5.41017364325691;5.79477650004684;6.04237760287688;5.68692117450
A2ML1_14639;5.4491082475774;5.36385961056552;5.17100665288747;5.6208513292
A4GALT_23662;8.36976136324847;7.80328185134206;8.55421636844331;7.22995624
A4GALT_9159;8.82297280814355;8.92285084244449;9.17229076620426;8.834796301
AAAS_61230;5.66491543991508;5.70809268946484;5.85667371223367;6.9757878359
AAAS_2894;5.89283524674721;5.6564212452558;5.66624190687793;6.090788421745
AAAS_18735;7.74349641324072;7.70023645818195;7.43741001836278;8.7585108190
AAAS 24237;6.41516988709249;6.40535808528583;6.82967565563798;7.0901431036
```



2. Set the dataset structure

- Column names are included in the first row: on
- File has row labels: on
- **Index of the column with row labels**: 1
- File has comment lines: off
- **Comment character**: #
- Switch rows and columns: on

For ASTERICS, individuals must be in rows.

Separator

○ , (comma) ● ; (semicolon) ○ \t (tab) ○ Space ○ Custom :

Quote mark for text

" (double quote)

Decimal mark

● . (dot) ○ , (comma)

Encoding

● Auto ○ UTF-8 ○ latin1

Missing values are encoded with these strings:

NA

Be patient: large files might take a long time to be uploaded and processed!

Raw file

```
;F476;F481;F560;F578;F688;F455;F485;F498;F738;F899;F900;F469;F557;F559;F6
A1CF_52154;5.82599980707194;5.91662162096354;5.62770189995087;5.974879117
A2M_40313;5.20290966528126;5.71662080795284;5.73529430299502;6.2799406003
A2M_10645;6.05043272137919;5.78608436449946;5.92397285829613;6.318404633
A2ML1_14639;5.09454305674891;4.89392730422894;5.06610396900029;5.10992723
A4GALT_23662;7.44499238746051;7.88769803161222;7.75157714142366;7.8547880
A4GALT_9159;8.80734586555401;8.74392533000205;8.55935004010872;8.2916779
AAAS_61230;5.82599980707194;5.60366933006277;6.19088825000998;6.69084861
AAAS_2894;5.59207508123505;5.29881778254516;5.87420980899597;5.9355008827
AAAS_18735;7.9656631095485;7.65139989082614;7.78415314320795;8.0945502888
```

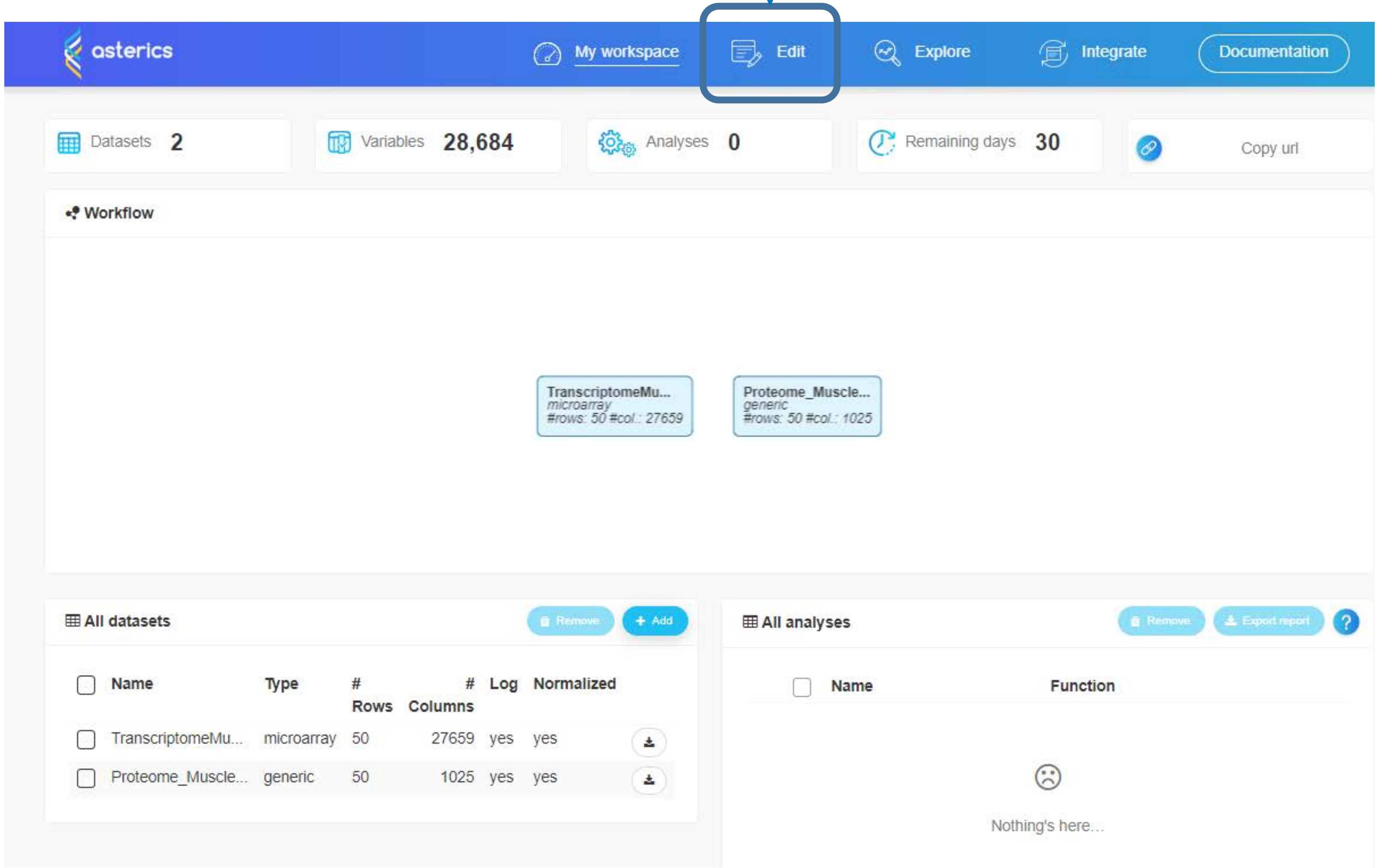
Dataset preview

	A1CF_52154 numeric	A2M_40313 numeric	A2M_10645 numeric	A2ML1_14639 numeric	A4GALT_23662 numeric
F476	5.826	5.2029	6.0504	5.0945	7.445
F481	5.9166	5.7166	5.7861	4.8939	7.8877
F560	5.6277	5.7353	5.924	5.0661	7.7516
F578	5.9749	6.2799	6.3184	5.1099	7.8548
F688	5.7889	6.434	6.5869	4.7516	8.0703
F455	5.5397	5.5826	5.8344	5.648	7.4885
F485	5.8868	6.1856	6.3453	5.661	7.7702
F498	5.6354	5.81	5.7552	5.8236	7.6827
F738	5.6221	5.1073	5.5711	5.6786	7.1388
F899	5.5986	5.5457	5.8591	5.0093	7.3409
F900	5.781	5.9127	6.1895	5.8663	7.7117
F469	6.1845	5.7959	5.7232	5.5926	8.0633

Switch rows and columns !

1 unique identifier per variable, so arrange the name to have the most relevant information !

Edit



Datasets 2

Variables 28,684

Analyses 0

Remaining days 30

Copy url

Workflow

TranscriptomeMu...
microarray
#rows: 50 #col.: 27659

Proteome_Muscle...
generic
#rows: 50 #col.: 1025

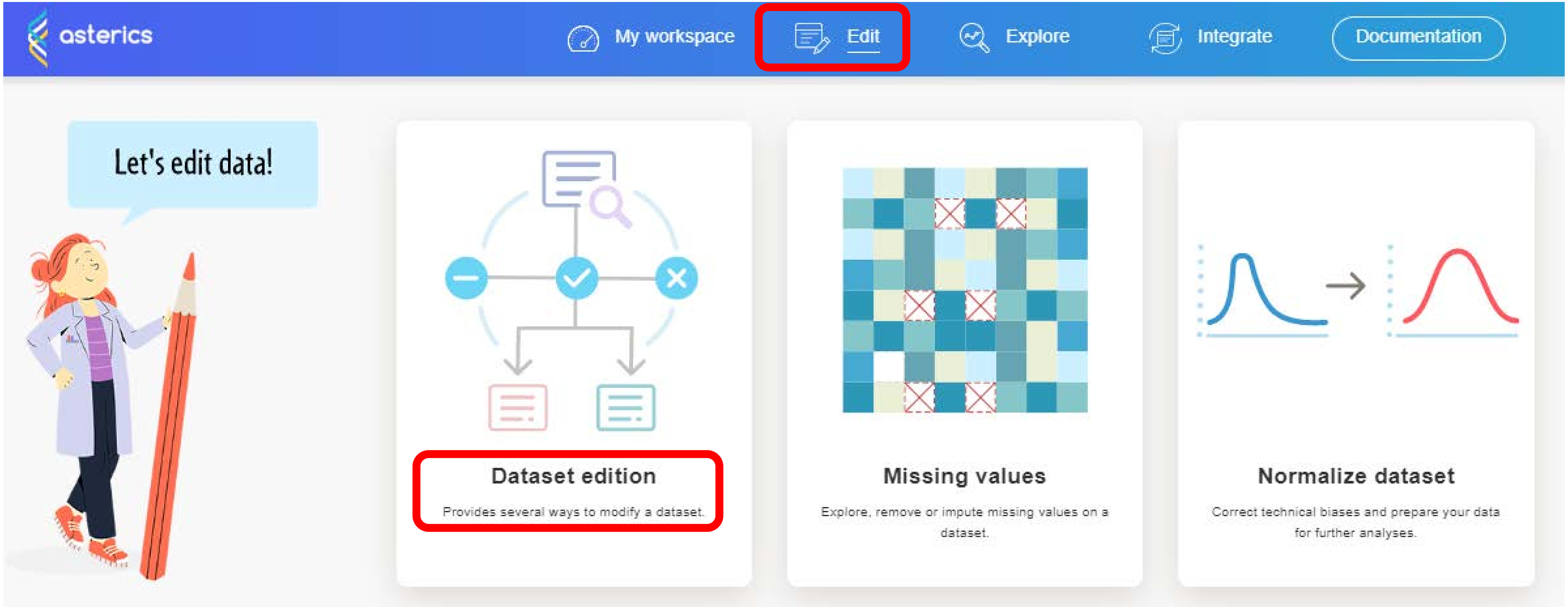
All datasets

☐	Name	Type	# Rows	# Columns	Log	Normalized
☐	TranscriptomeMu...	microarray	50	27659	yes	yes
☐	Proteome_Muscle...	generic	50	1025	yes	yes

All analyses

☐	Name	Function

Nothing's here...



Dataset edition

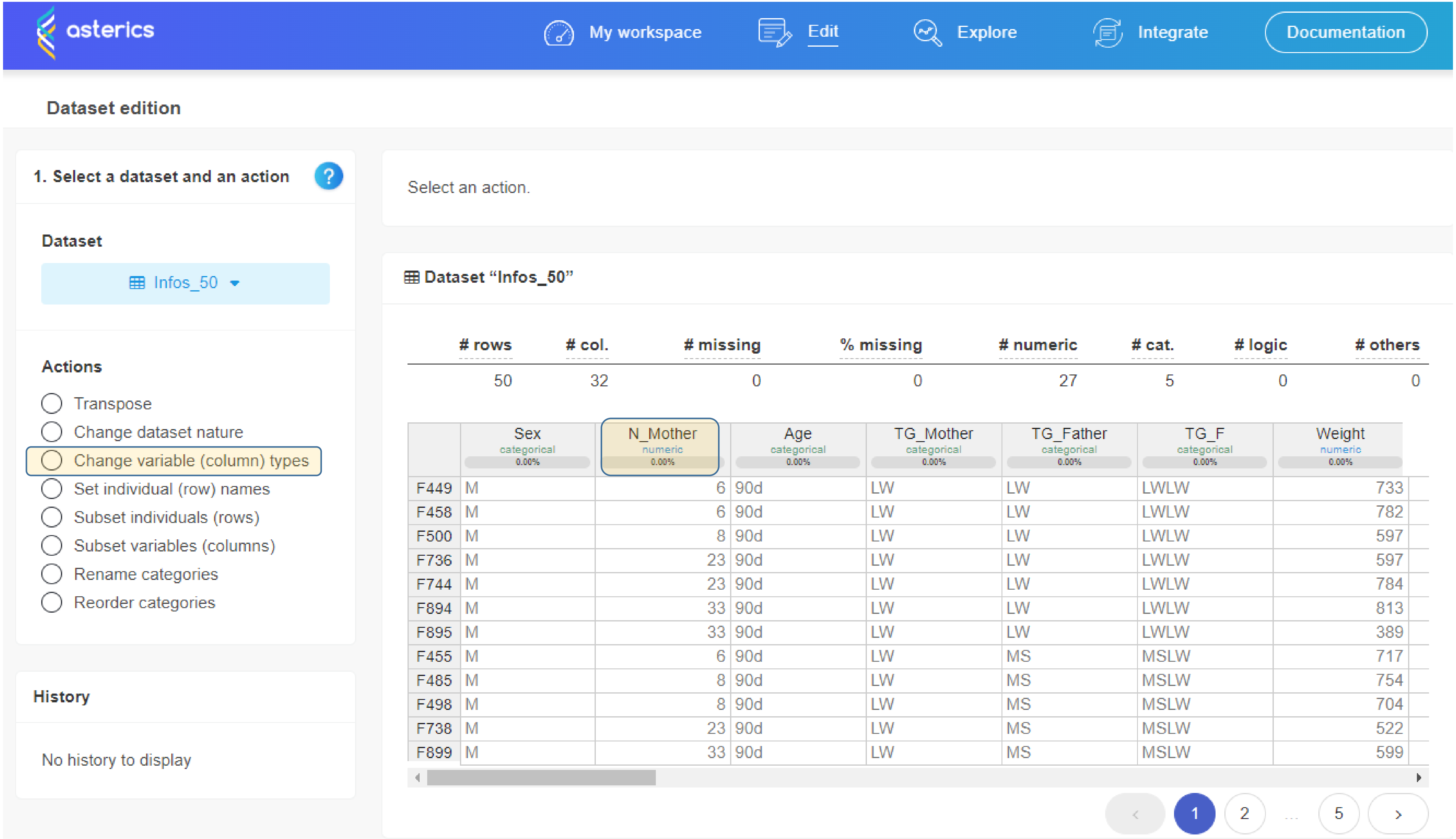
Provides several ways to modify a dataset.

Missing values

Explore, remove or impute missing values on a dataset.

Normalize dataset

Correct technical biases and prepare your data for further analyses.



Dataset edition

1. Select a dataset and an action

Dataset

Infos_50

Actions

○ Transpose
○ Change dataset nature
○ Change variable (column) types
○ Set individual (row) names
○ Subset individuals (rows)
○ Subset variables (columns)
○ Rename categories
○ Reorder categories

History

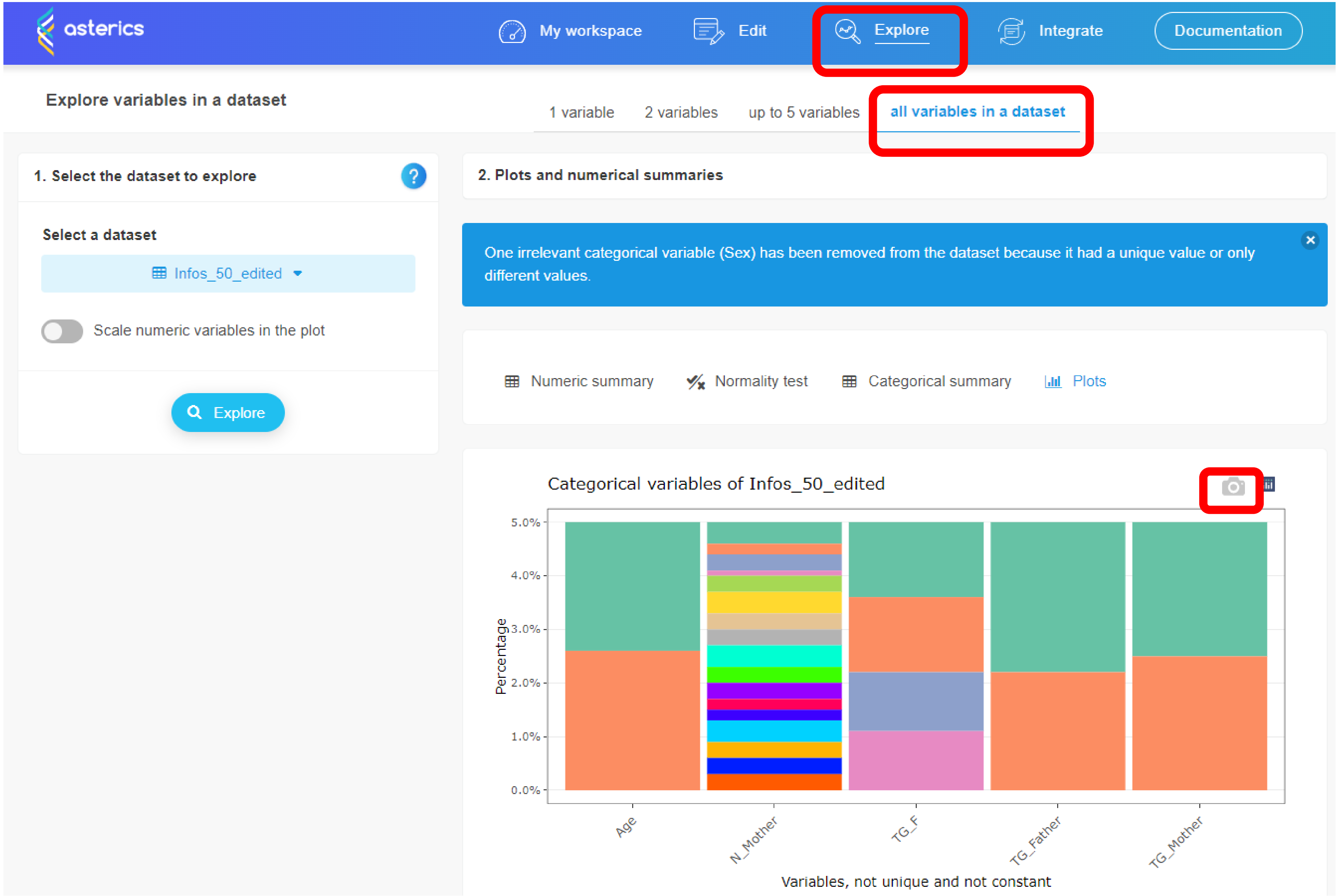
No history to display

Select an action.

Dataset "Infos_50"

# rows	# col.	# missing	% missing	# numeric	# cat.	# logic	# others
50	32	0	0	27	5	0	0

	Sex categorical 0.00%	N_Mother numeric 0.00%	Age categorical 0.00%	TG_Mother categorical 0.00%	TG_Father categorical 0.00%	TG_F categorical 0.00%	Weight numeric 0.00%
F449	M	6	90d	LW	LW	LWLW	733
F458	M	6	90d	LW	LW	LWLW	782
F500	M	8	90d	LW	LW	LWLW	597
F736	M	23	90d	LW	LW	LWLW	597
F744	M	23	90d	LW	LW	LWLW	784
F894	M	33	90d	LW	LW	LWLW	813
F895	M	33	90d	LW	LW	LWLW	389
F455	M	6	90d	LW	MS	MSLW	717
F485	M	8	90d	LW	MS	MSLW	754
F498	M	8	90d	LW	MS	MSLW	704
F738	M	23	90d	LW	MS	MSLW	522
F899	M	33	90d	LW	MS	MSLW	599



Explore variables in a dataset

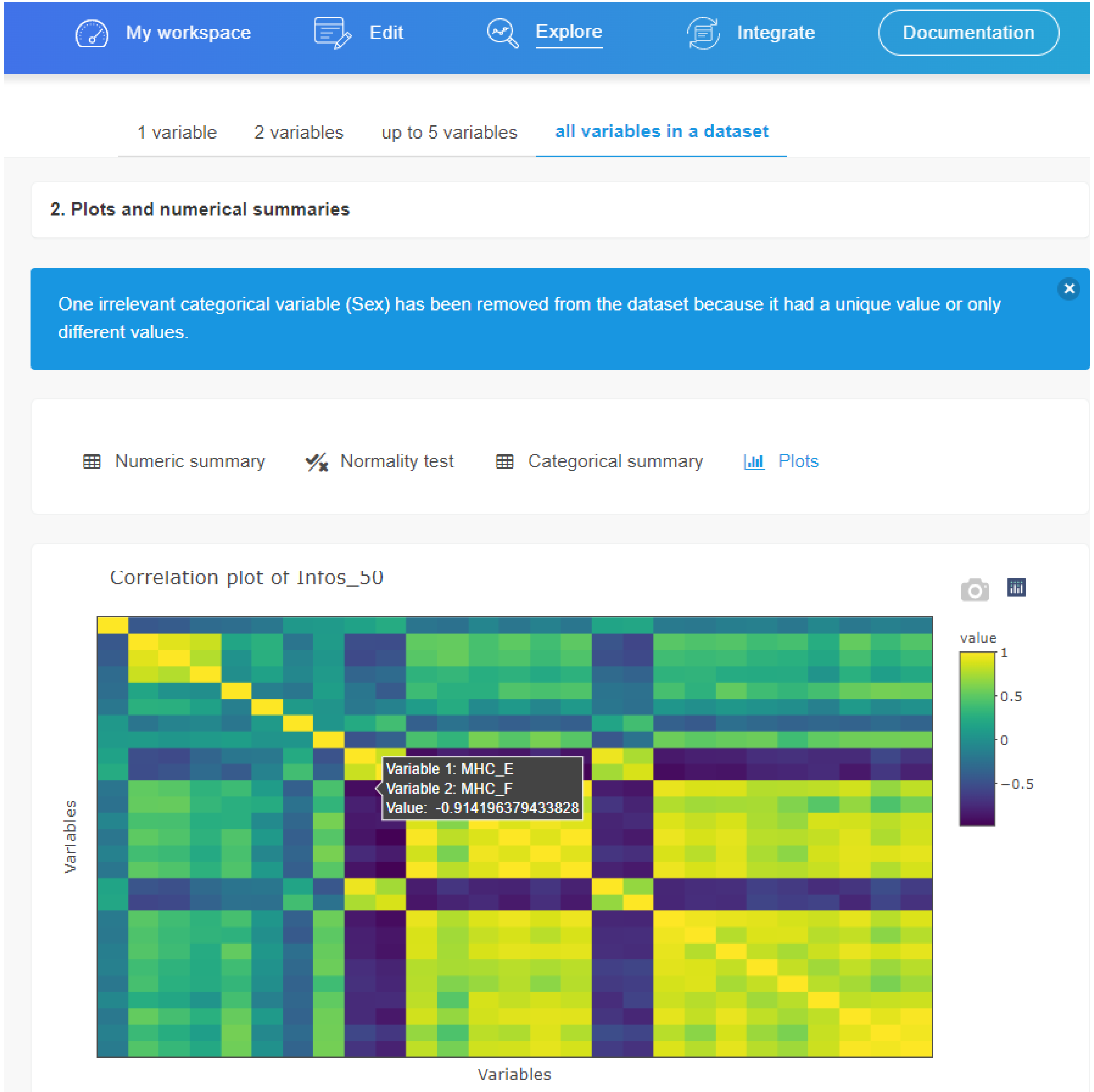
1. Select the dataset to explore

Select a dataset: Infos_50_edited

Scale numeric variables in the plot

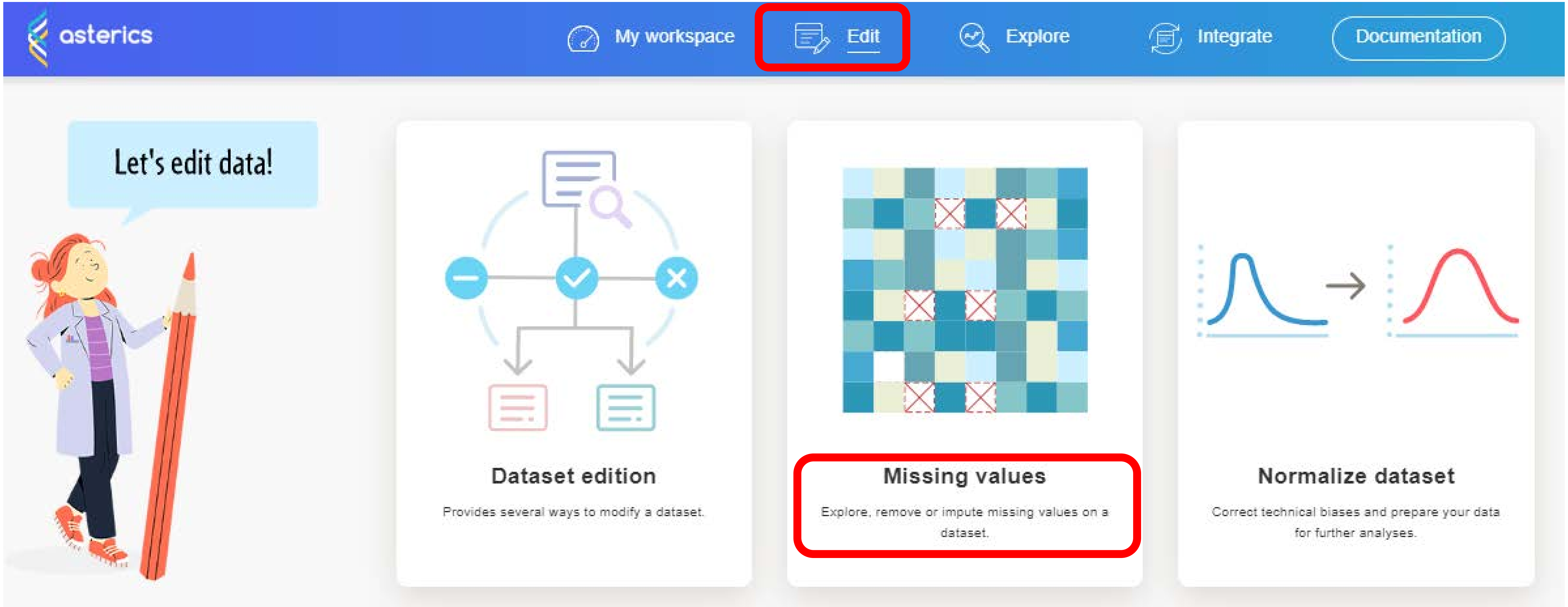
2. Plots and numerical summaries

One irrelevant categorical variable (Sex) has been removed from the dataset because it had a unique value or only different values.



2. Plots and numerical summaries

One irrelevant categorical variable (Sex) has been removed from the dataset because it had a unique value or only different values.



Dataset edition

Provides several ways to modify a dataset.

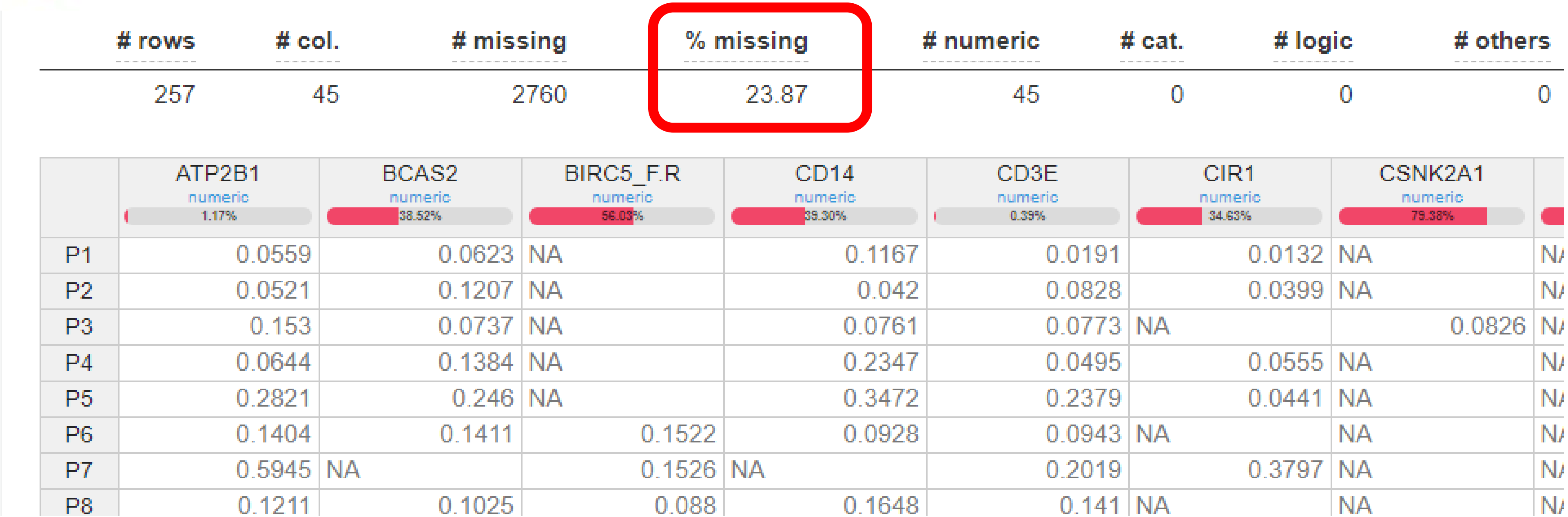
Missing values

Explore, remove or impute missing values on a dataset.

Normalize dataset

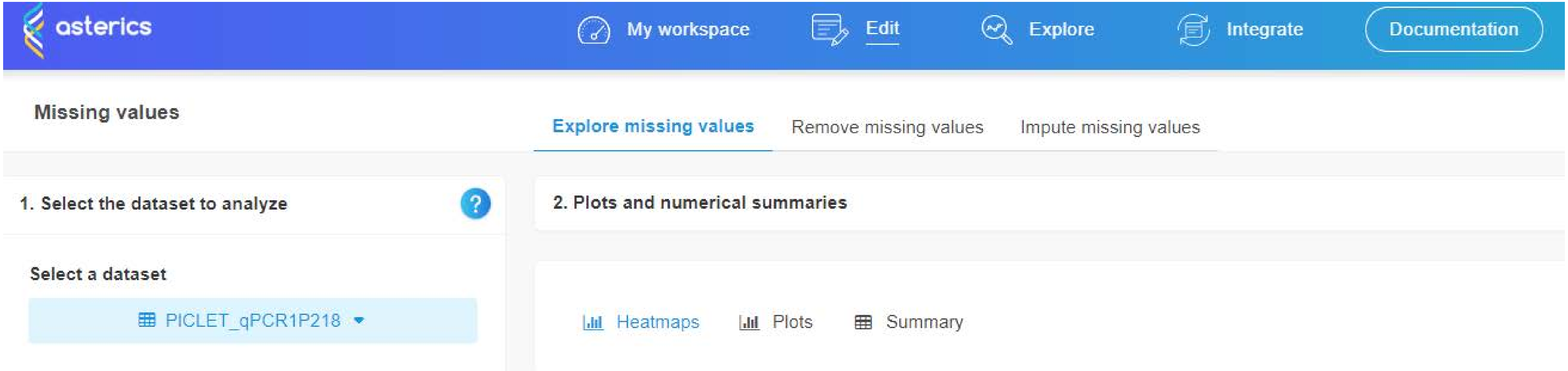
Correct technical biases and prepare your data for further analyses.

qPCR data...



# rows	# col.	# missing	% missing	# numeric	# cat.	# logic	# others
257	45	2760	23.87	45	0	0	0

	ATP2B1 numeric 1.17%	BCAS2 numeric 38.52%	BIRC5_F.R numeric 56.03%	CD14 numeric 39.30%	CD3E numeric 0.39%	CIR1 numeric 34.63%	CSNK2A1 numeric 79.38%	
P1	0.0559	0.0623	NA	0.1167	0.0191	0.0132	NA	N/
P2	0.0521	0.1207	NA	0.042	0.0828	0.0399	NA	N/
P3	0.153	0.0737	NA	0.0761	0.0773	NA	0.0826	N/
P4	0.0644	0.1384	NA	0.2347	0.0495	0.0555	NA	N/
P5	0.2821	0.246	NA	0.3472	0.2379	0.0441	NA	N/
P6	0.1404	0.1411	0.1522	0.0928	0.0943	NA	NA	N/
P7	0.5945	NA	0.1526	NA	0.2019	0.3797	NA	N/
P8	0.1211	0.1025	0.088	0.1648	0.141	NA	NA	N/



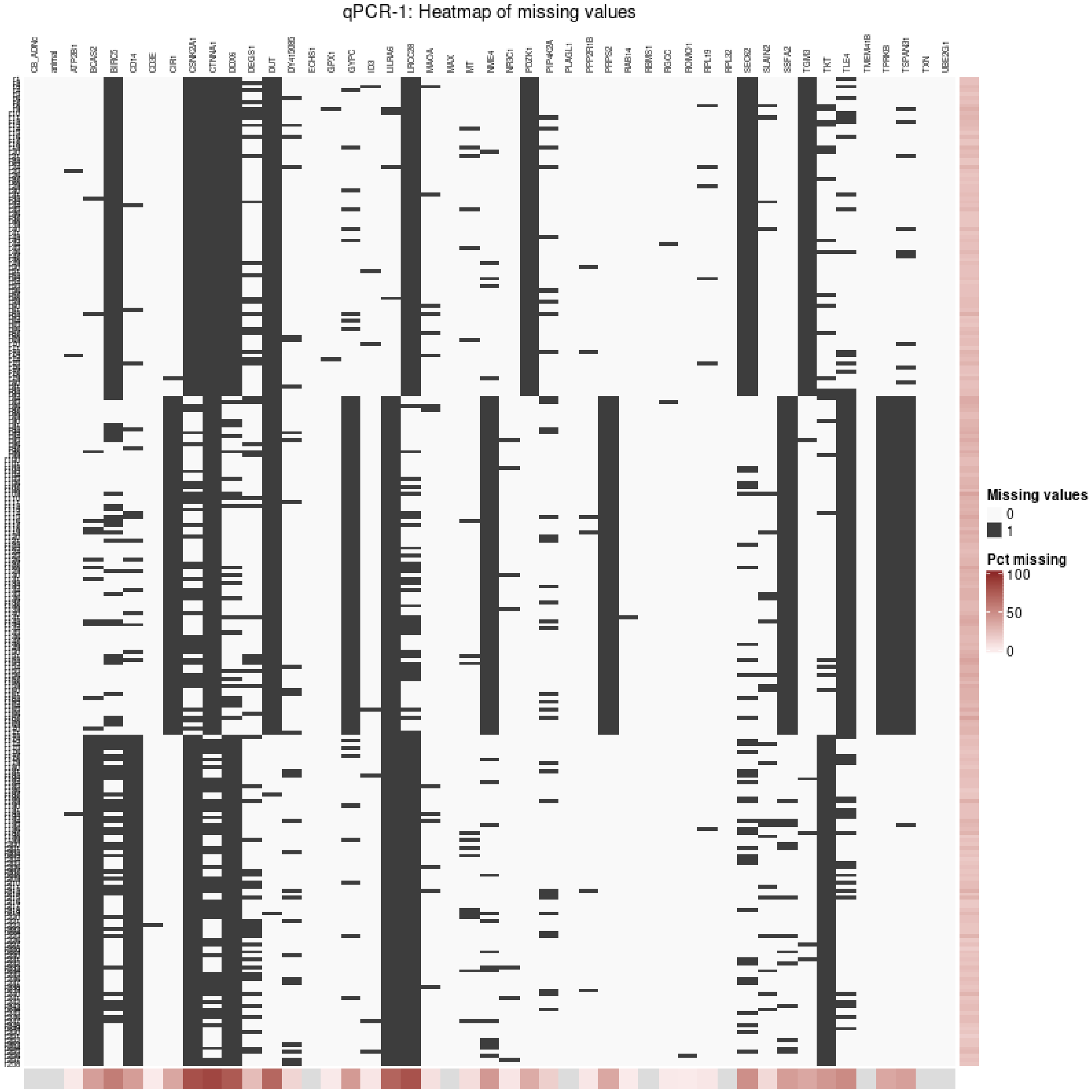
Missing values

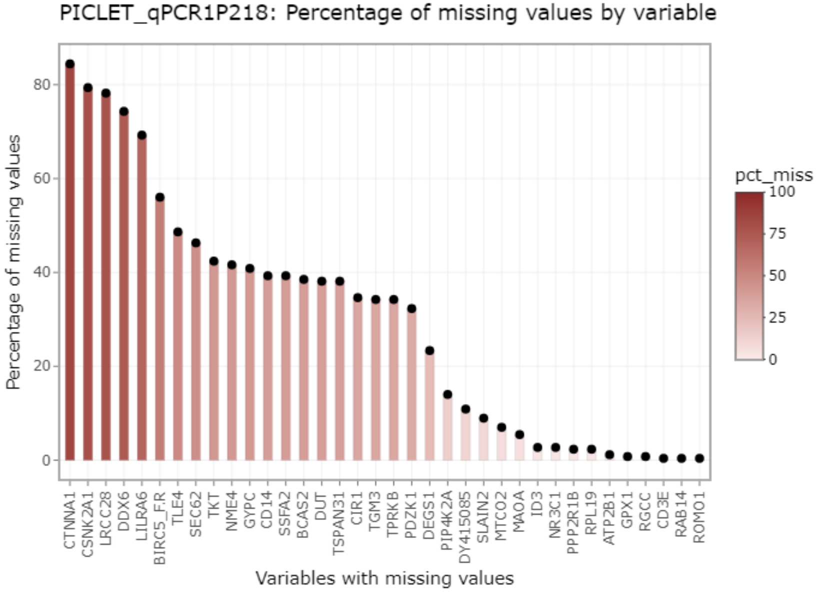
1. Select the dataset to analyze

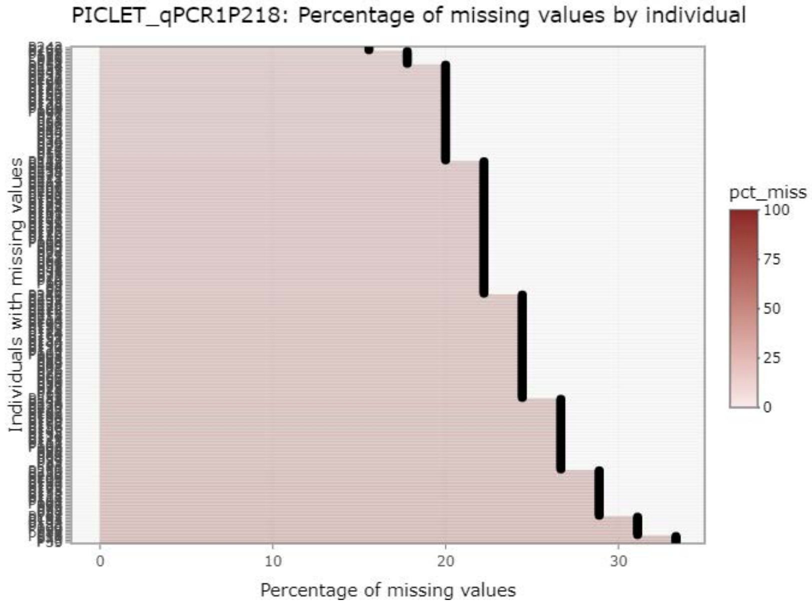
Select a dataset

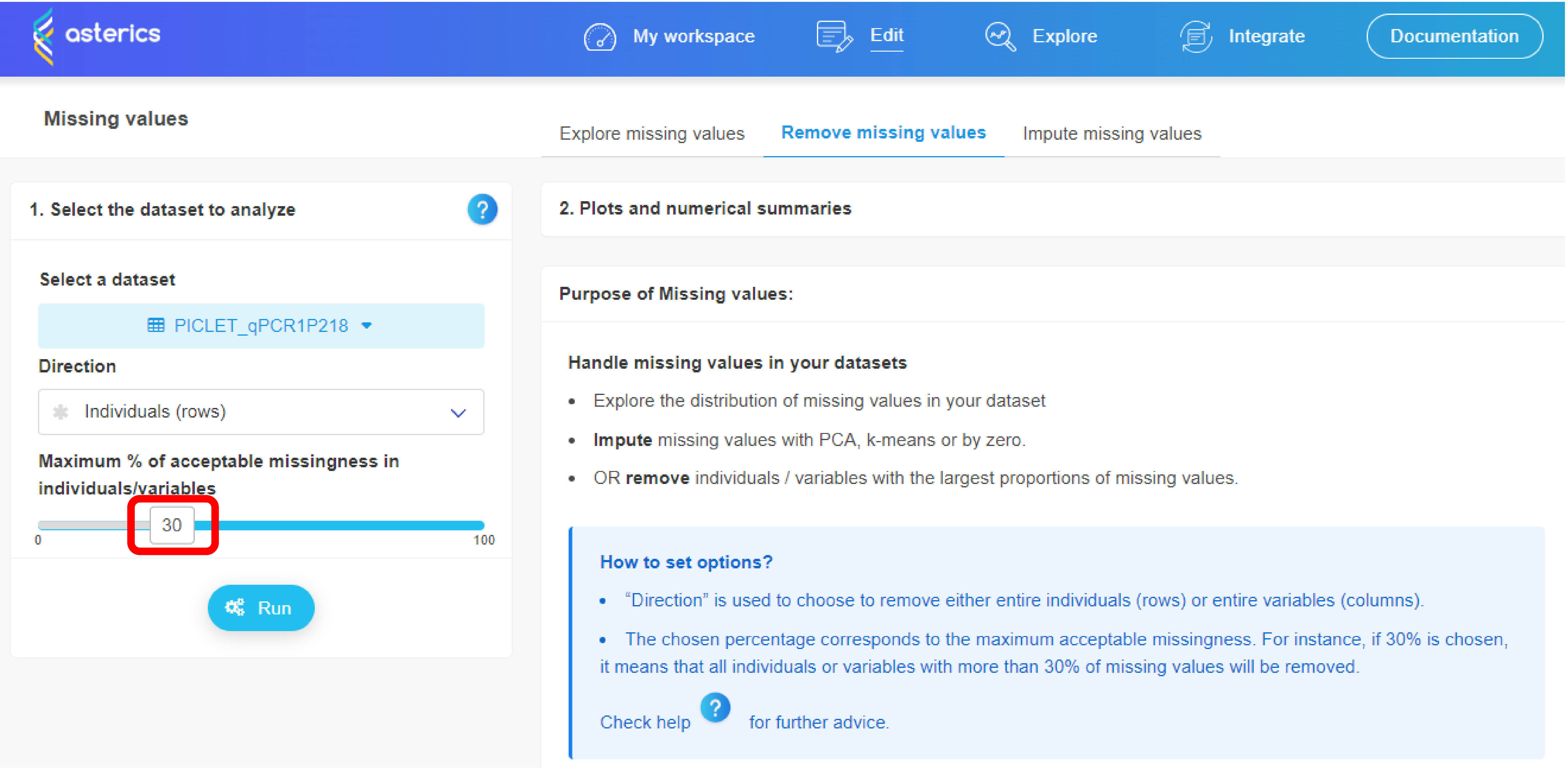
PICLET_qPCR1P218

2. Plots and numerical summaries


qPCR-1: Heatmap of missing values
CB_ADNc
animal
ATP2B1
BCAS2
BIRC5
CD14
CD3E
CIR1
CSNK2A1
CTNNA1
DDX6
DEGS1
DUT
DY415085
ECHS1
GPX1
GYPC
ID3
LILRA6
LRCC28
MAOA
MAX
MT
NME4
NR3C1
PDZK1
PIP4K2A
PLAGL1
PPP2R1B
PRPS2
RAB14
RBMS1
RGCC
ROMO1
RPL19
RPL32
SEC62
SLAIN2
SSFA2
TGM3
TKT
TLE4
TMEM41B
TPRKB
TSPAN31
TXN
UBE2G1
Missing values
0
1
Pct missing
100
50
0


PICLET_qPCR1P218: Percentage of missing values by variable
Percentage of missing values
80
60
40
20
0
CTNNA1
CSNK2A1
LRCC28
DDX6
LILRA6
BIRC5_FR
TLE4
SEC62
TKT
NME4
GYPC
CD14
SSFA2
BCAS2
DUT
TSPAN31
CIR1
TGM3
TPRKB
PDZK1
DEGS1
PIP4K2A
DY415085
SLAIN2
MTCO2
MAOA
ID3
NR3C1
PPP2R1B
RPL19
ATP2B1
GPX1
RGCC
CD3E
RAB14
ROMO1
Variables with missing values
pct_miss
100
75
50
25
0


PICLET_qPCR1P218: Percentage of missing values by individual
Individuals with missing values
0
10
20
30
Percentage of missing values
pct_miss
100
75
50
25
0



1. Select the dataset to analyze

Select a dataset

PICLET_qPCR1P218

Direction

Individuals (rows)

Maximum % of acceptable missingness in individuals/variables

0 — 30 — 100

2. Plots and numerical summaries

Purpose of Missing values:

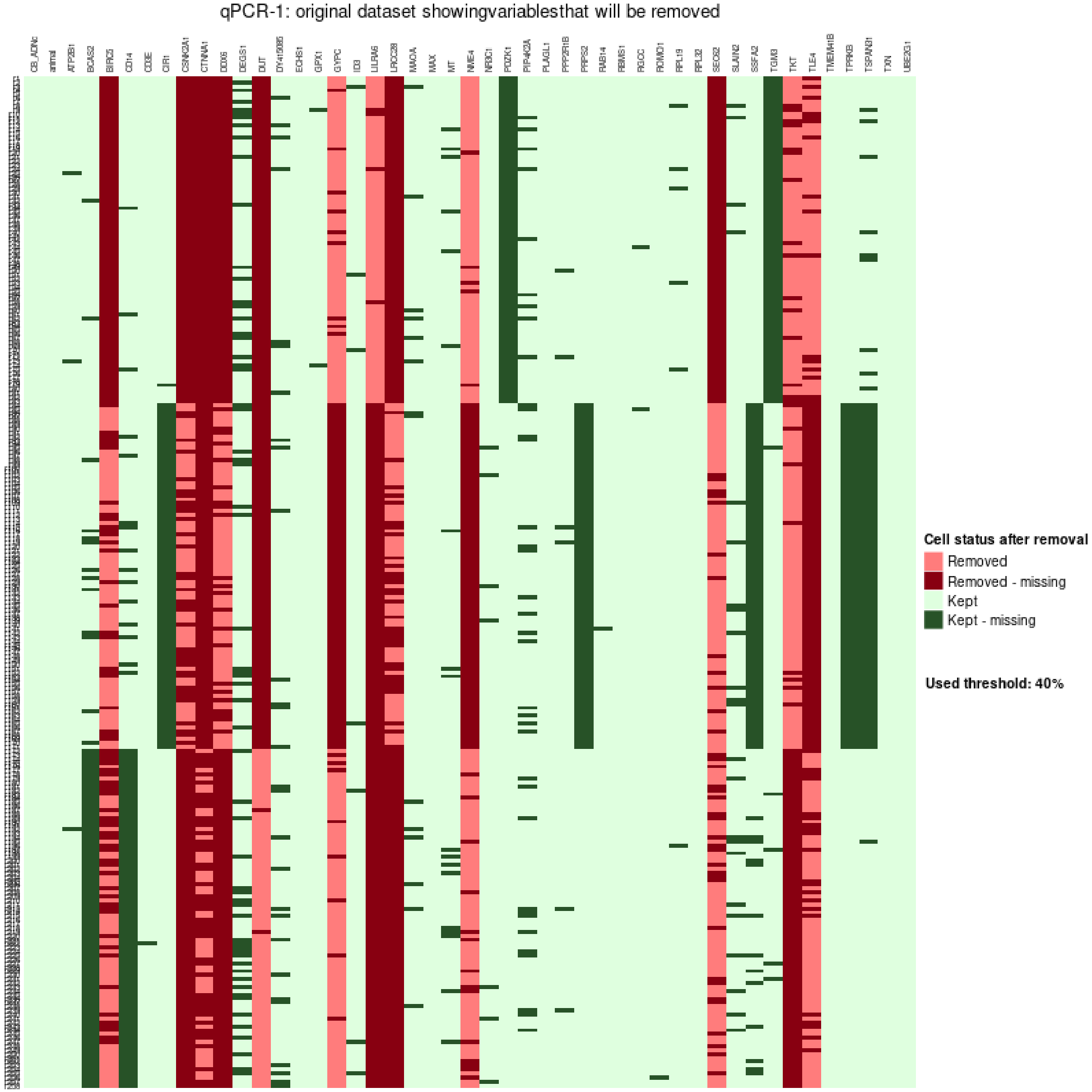
Handle missing values in your datasets

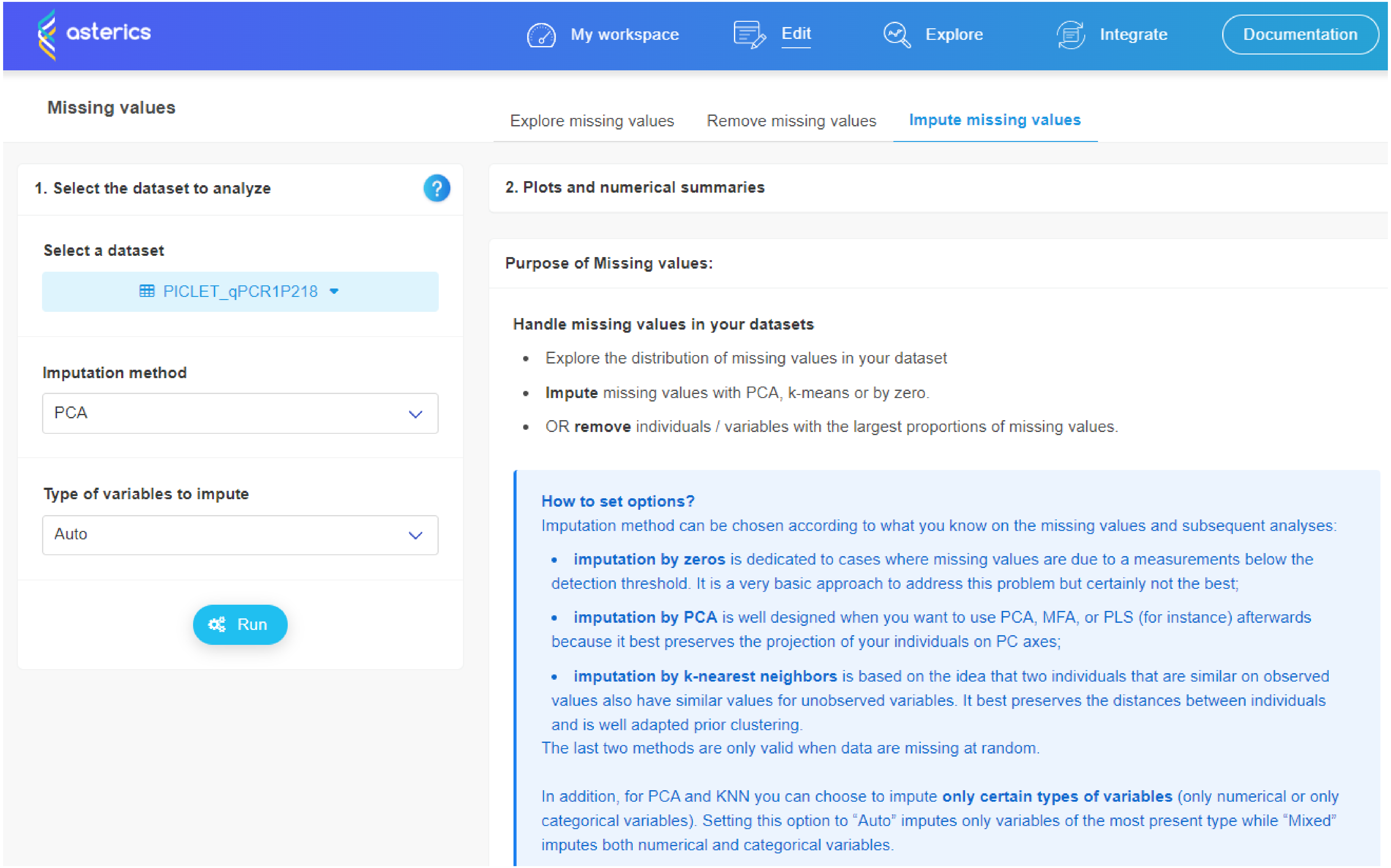
- Explore the distribution of missing values in your dataset
- **Impute** missing values with PCA, k-means or by zero.
- OR **remove** individuals / variables with the largest proportions of missing values.

How to set options?

- "Direction" is used to choose to remove either entire individuals (rows) or entire variables (columns).
- The chosen percentage corresponds to the maximum acceptable missingness. For instance, if 30% is chosen, it means that all individuals or variables with more than 30% of missing values will be removed.

Check help ? for further advice.


qPCR-1: original dataset showingvariablesthat will be removed



1. Select the dataset to analyze

Select a dataset

PICLET_qPCR1P218

Imputation method

PCA

Type of variables to impute

Auto

2. Plots and numerical summaries

Purpose of Missing values:

Handle missing values in your datasets

- Explore the distribution of missing values in your dataset
- **Impute** missing values with PCA, k-means or by zero.
- OR **remove** individuals / variables with the largest proportions of missing values.

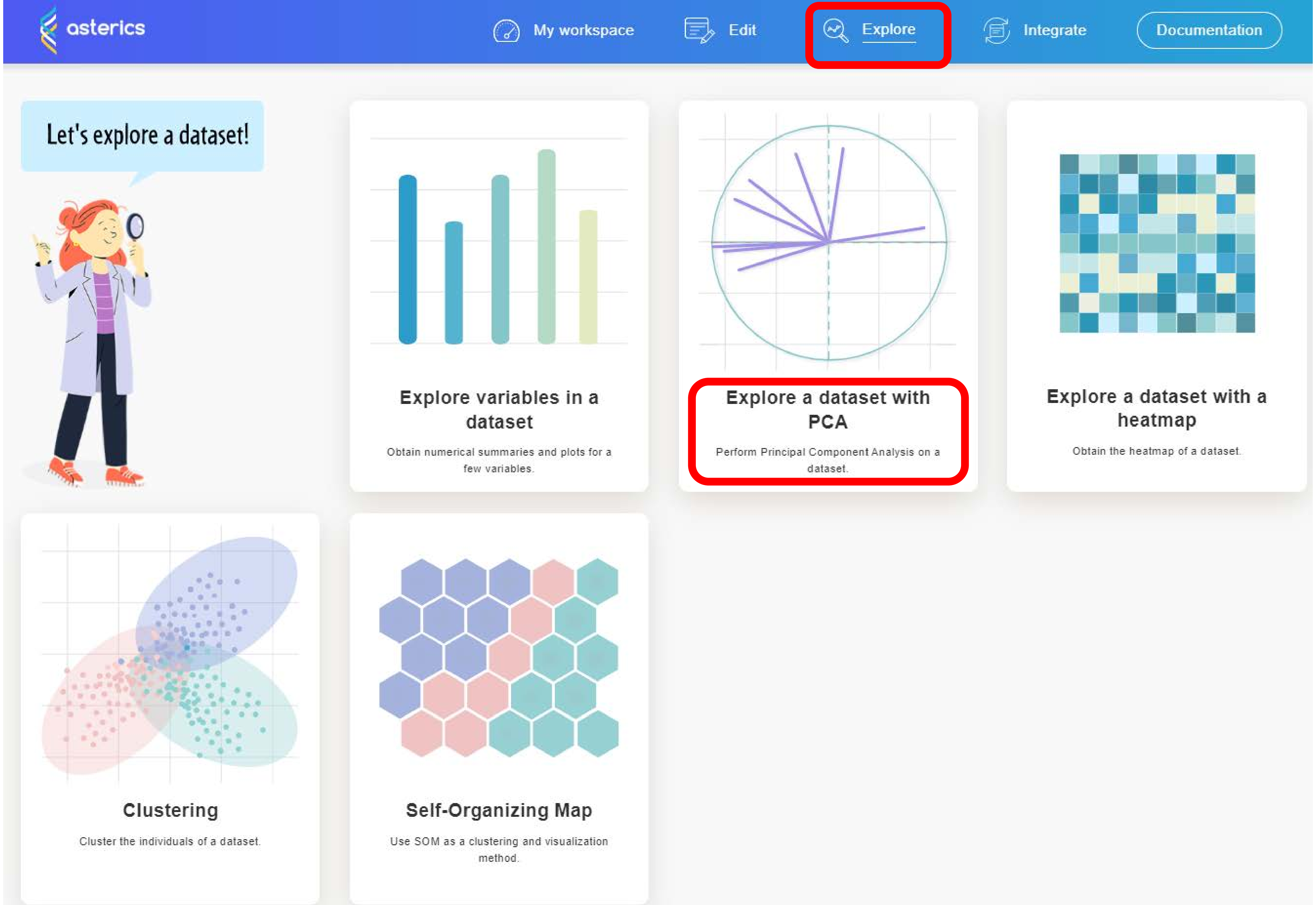
How to set options?

Imputation method can be chosen according to what you know on the missing values and subsequent analyses:

- **imputation by zeros** is dedicated to cases where missing values are due to a measurements below the detection threshold. It is a very basic approach to address this problem but certainly not the best;
- **imputation by PCA** is well designed when you want to use PCA, MFA, or PLS (for instance) afterwards because it best preserves the projection of your individuals on PC axes;
- **imputation by k-nearest neighbors** is based on the idea that two individuals that are similar on observed values also have similar values for unobserved variables. It best preserves the distances between individuals and is well adapted prior clustering.

The last two methods are only valid when data are missing at random.

In addition, for PCA and KNN you can choose to impute **only certain types of variables** (only numerical or only categorical variables). Setting this option to “Auto” imputes only variables of the most present type while “Mixed” imputes both numerical and categorical variables.



Explore variables in a dataset

Obtain numerical summaries and plots for a few variables.

Explore a dataset with PCA

Perform Principal Component Analysis on a dataset.

Explore a dataset with a heatmap

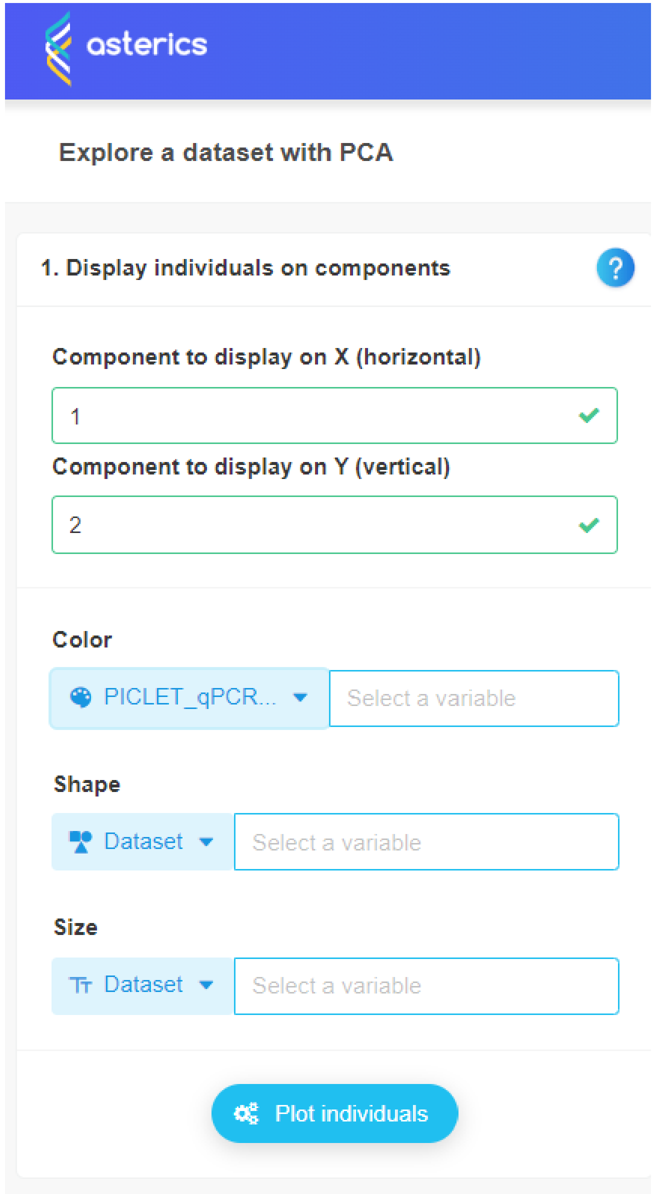
Obtain the heatmap of a dataset.

Clustering

Cluster the individuals of a dataset.

Self-Organizing Map

Use SOM as a clustering and visualization method.



Explore a dataset with PCA

1. Display individuals on components

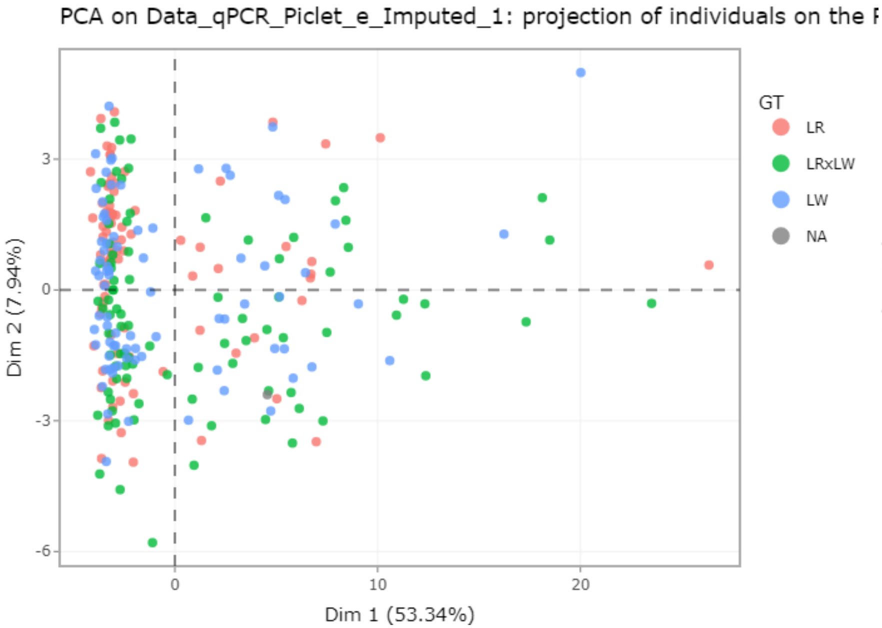
Component to display on X (horizontal): 1

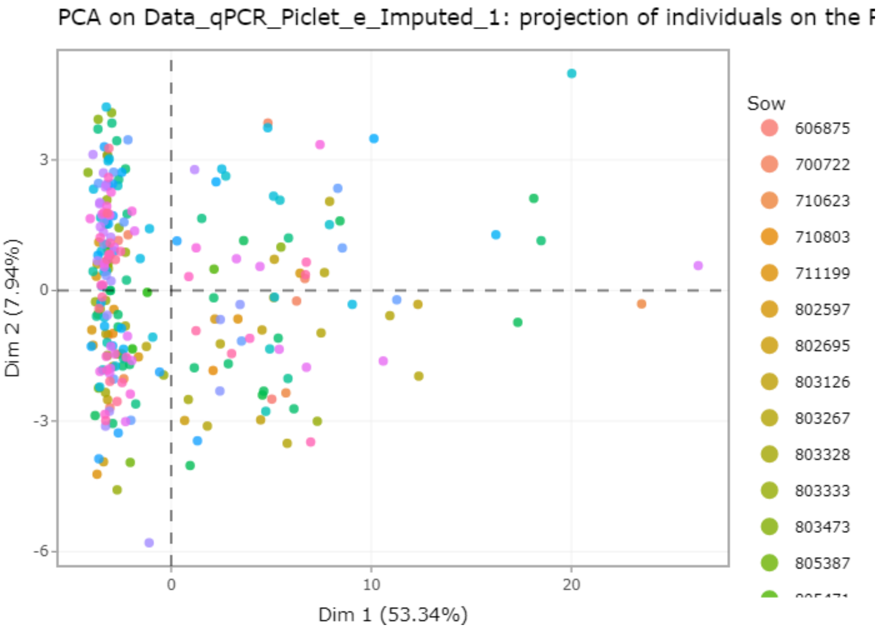
Component to display on Y (vertical): 2

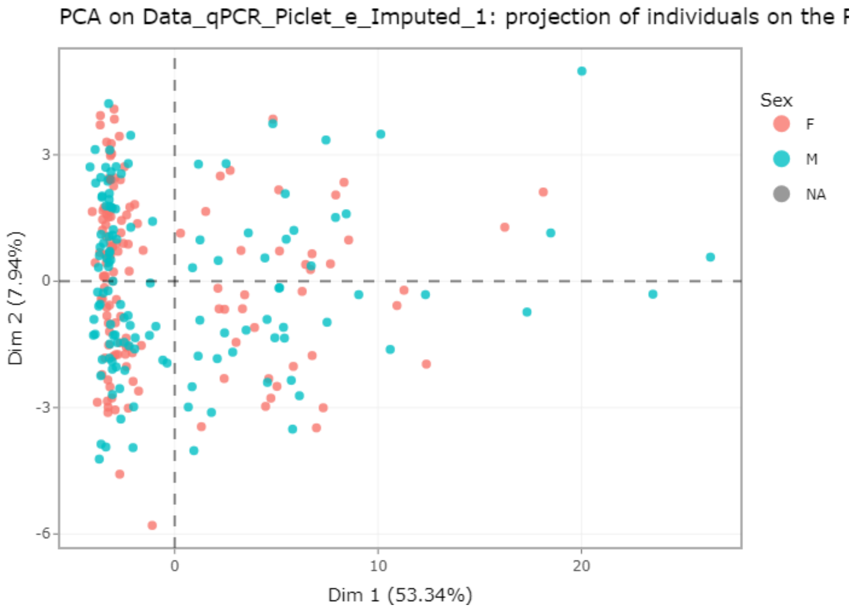
Color: PICLET_qPCR... | Select a variable

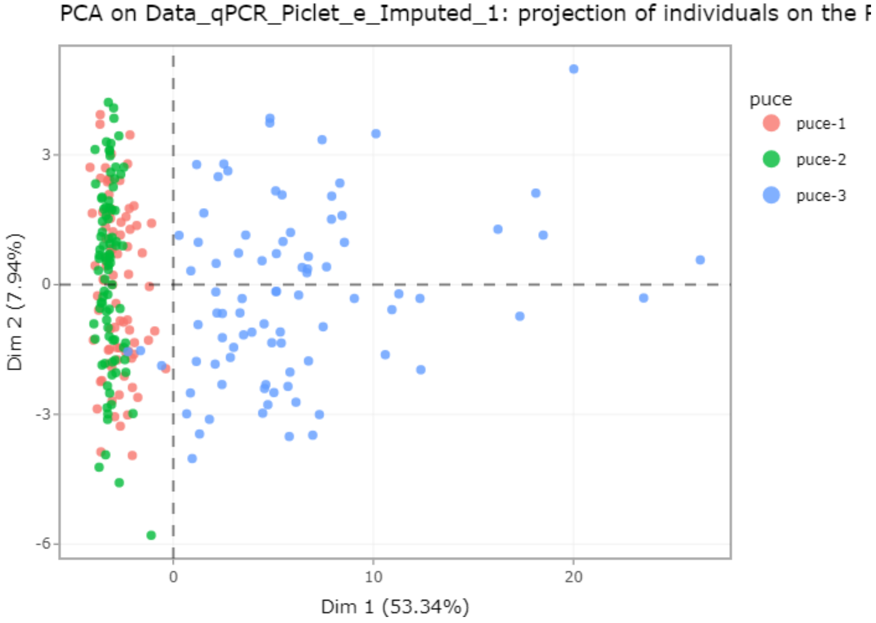
Shape: Dataset | Select a variable

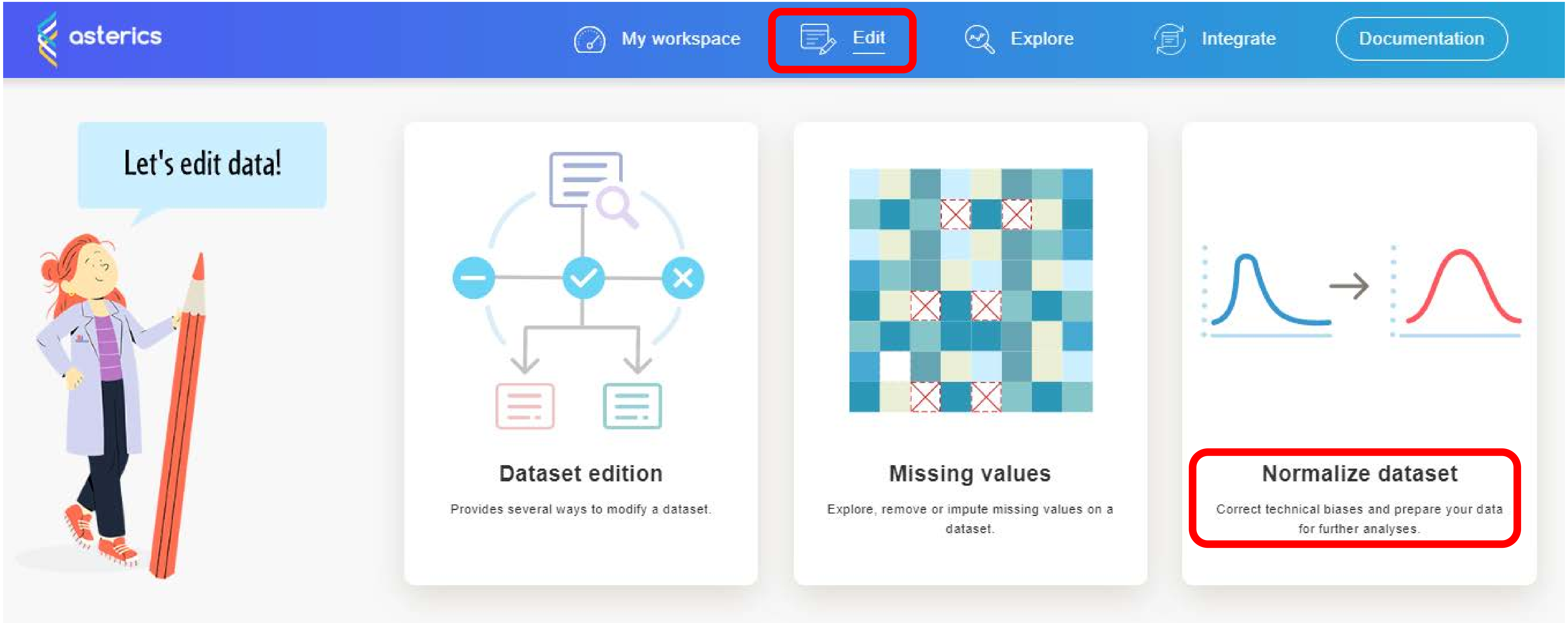
Size: Dataset | Select a variable











Dataset edition

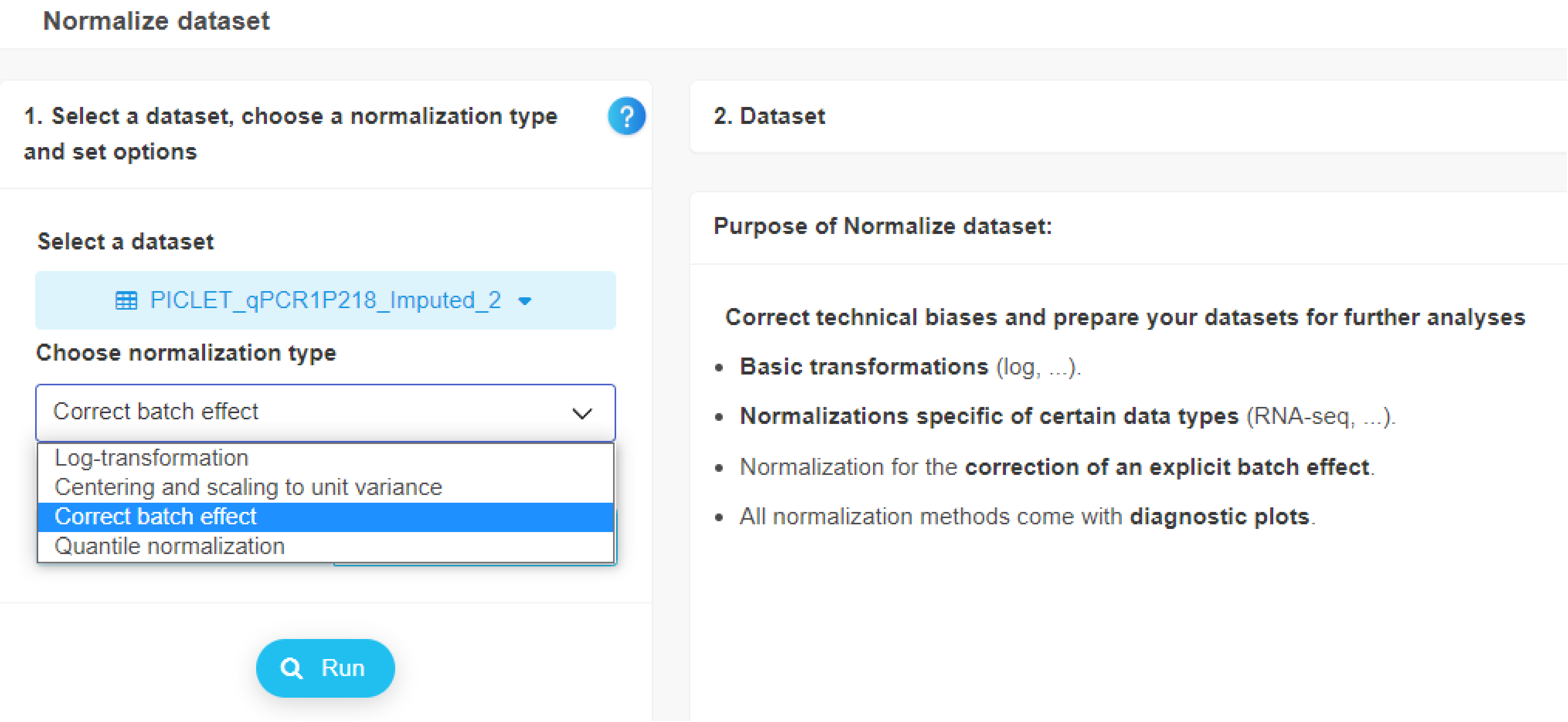
Provides several ways to modify a dataset.

Missing values

Explore, remove or impute missing values on a dataset.

Normalize dataset

Correct technical biases and prepare your data for further analyses.



Normalize dataset

1. Select a dataset, choose a normalization type and set options

Select a dataset

PICLET_qPCR1P218_Imputed_2

Choose normalization type

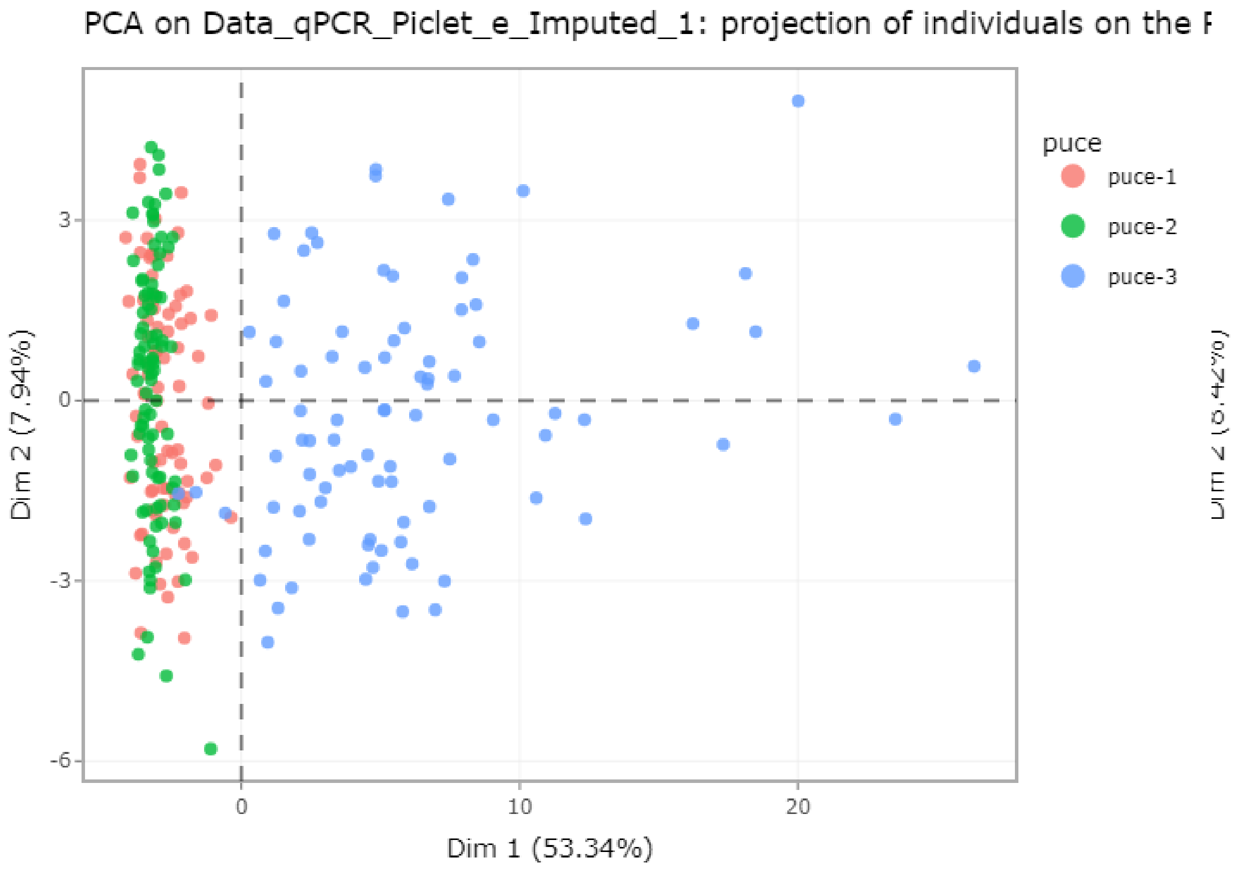
Correct batch effect

2. Dataset

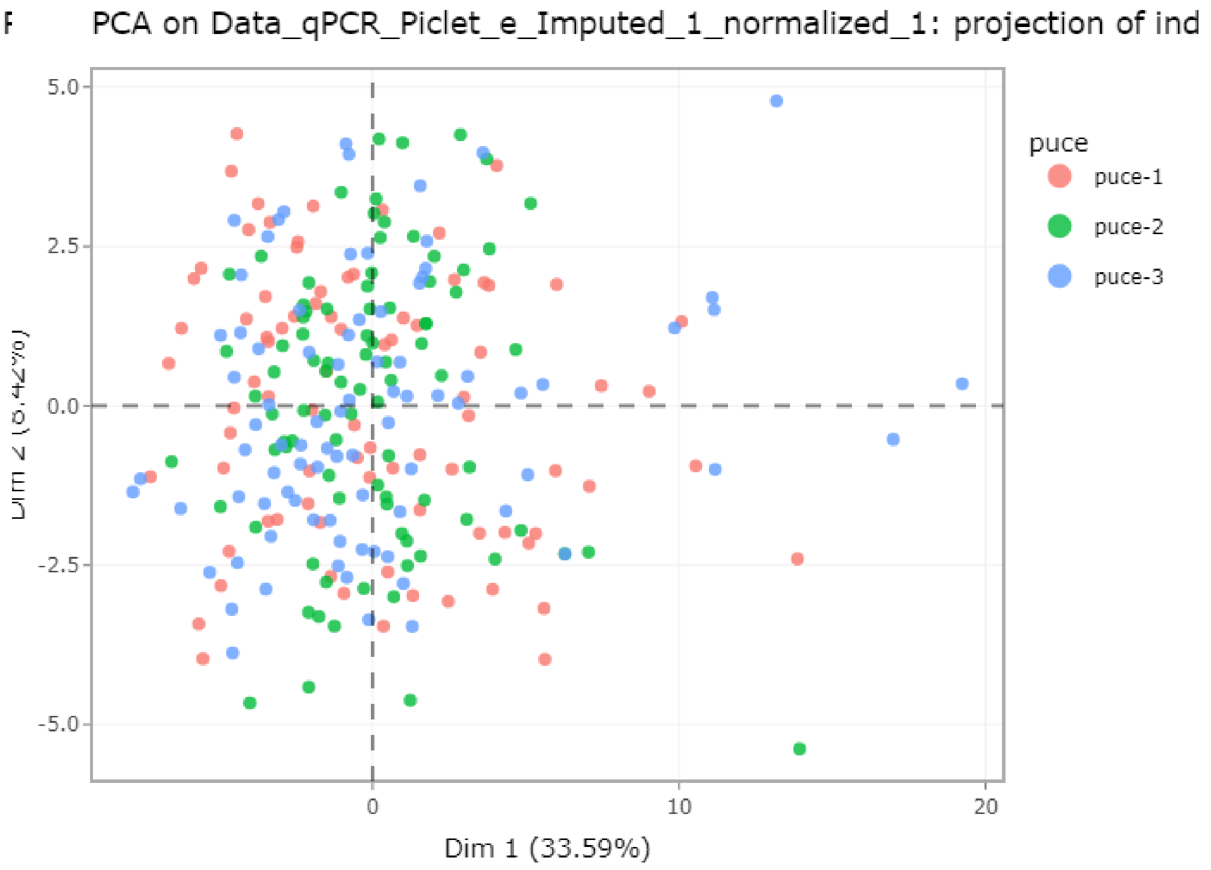
Purpose of Normalize dataset:

Correct technical biases and prepare your datasets for further analyses

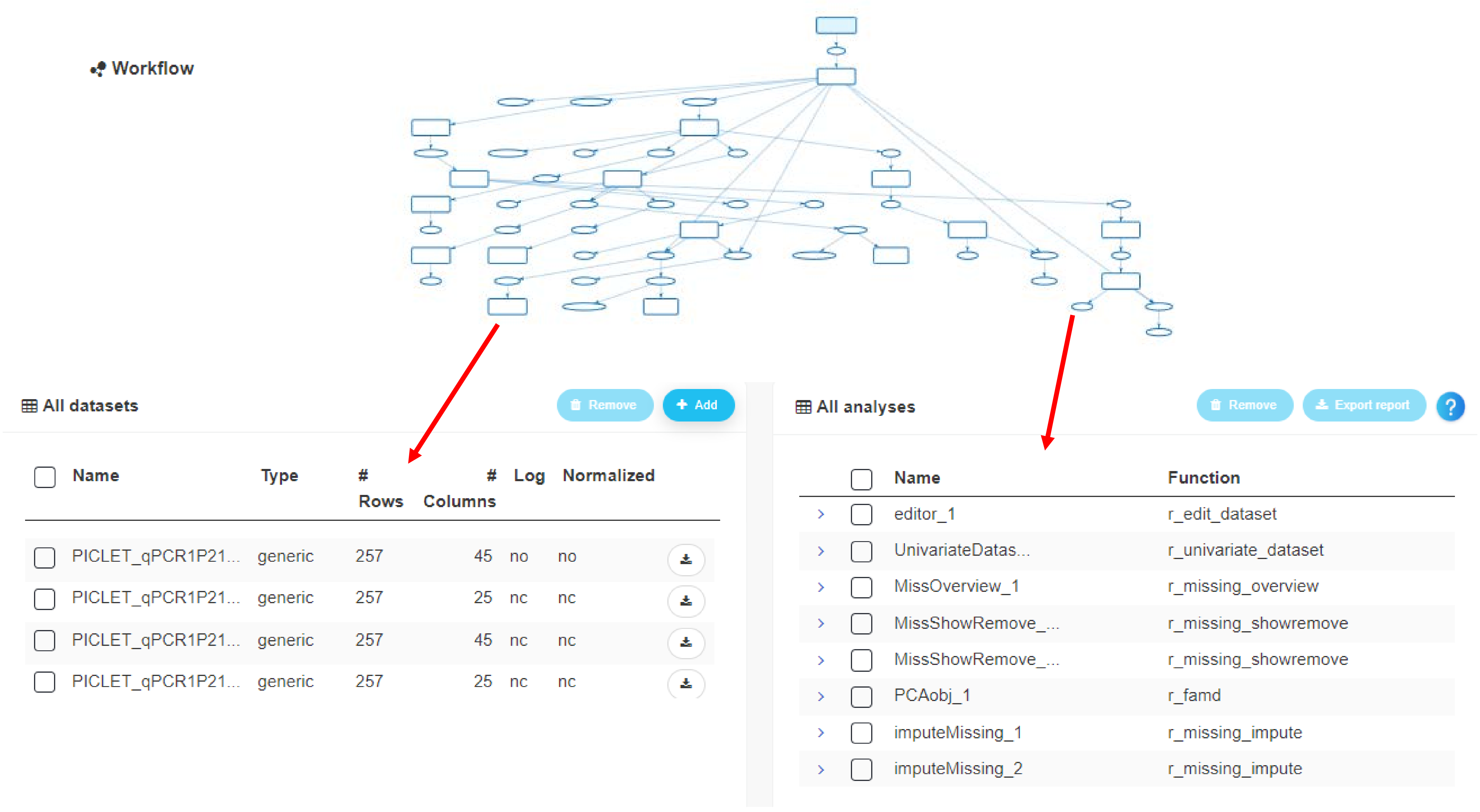
- **Basic transformations** (log, ...).
- **Normalizations specific of certain data types** (RNA-seq, ...).
- Normalization for the **correction of an explicit batch effect**.
- All normalization methods come with **diagnostic plots**.



PCA on Data_qPCR_Piclet_e_Imputed_1: projection of individuals on the



PCA on Data_qPCR_Piclet_e_Imputed_1_normalized_1: projection of ind



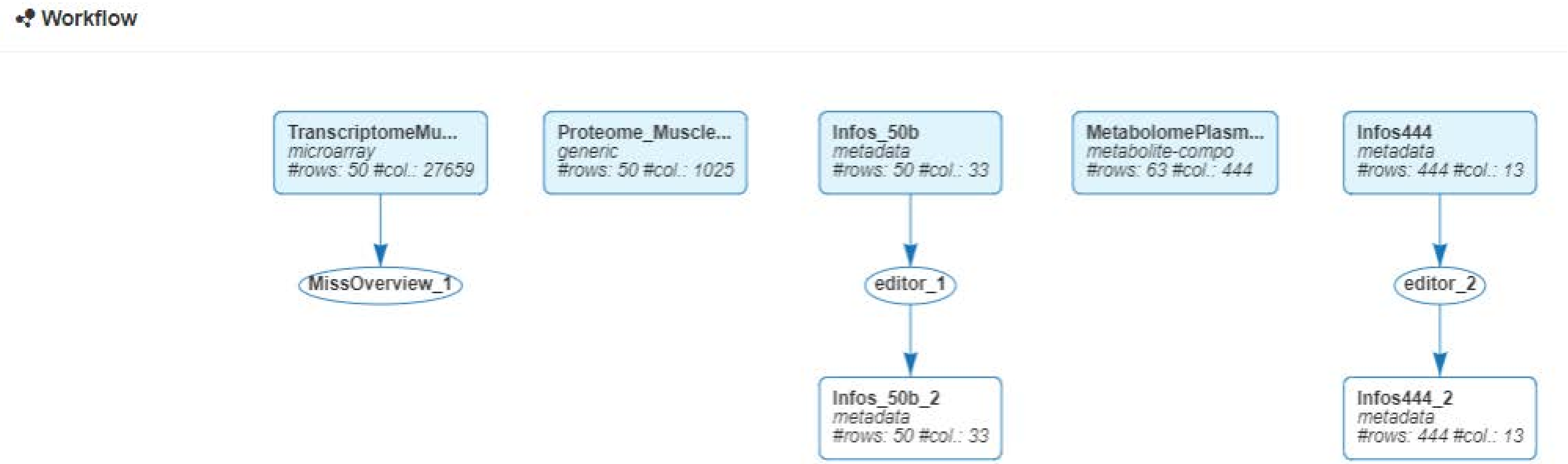
Workflow

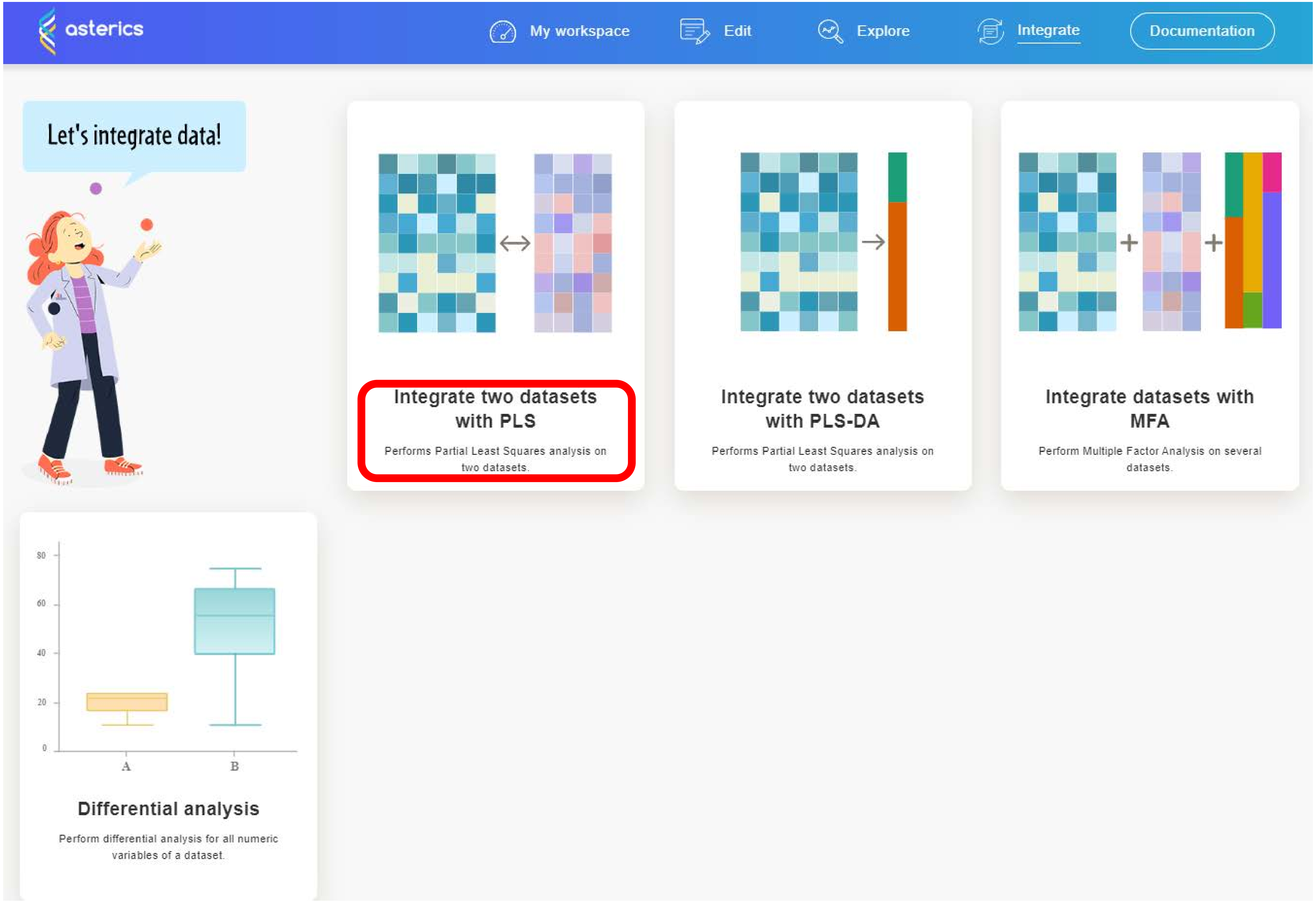
All datasets

☐	Name	Type	# Rows	# Columns	Log	Normalized	
☐	PICLET_qPCR1P21...	generic	257	45	no	no	⤓
☐	PICLET_qPCR1P21...	generic	257	25	nc	nc	⤓
☐	PICLET_qPCR1P21...	generic	257	45	nc	nc	⤓
☐	PICLET_qPCR1P21...	generic	257	25	nc	nc	⤓

All analyses

	☐	Name	Function
›	☐	editor_1	r_edit_dataset
›	☐	UnivariateDatas...	r_univariate_dataset
›	☐	MissOverview_1	r_missing_overview
›	☐	MissShowRemove_...	r_missing_showremove
›	☐	MissShowRemove_...	r_missing_showremove
›	☐	PCAobj_1	r_famd
›	☐	imputeMissing_1	r_missing_impute
›	☐	imputeMissing_2	r_missing_impute


Workflow
TranscriptomeMu...
microarray
#rows: 50 #col.: 27659
Proteome_Muscle...
generic
#rows: 50 #col.: 1025
Infos_50b
metadata
#rows: 50 #col.: 33
MetabolomePlasm...
metabolite-compo
#rows: 63 #col.: 444
Infos444
metadata
#rows: 444 #col.: 13
MissOverview_1
editor_1
editor_2
Infos_50b_2
metadata
#rows: 50 #col.: 33
Infos444_2
metadata
#rows: 444 #col.: 13



Integrate two datasets with PLS

Performs Partial Least Squares analysis on two datasets.

Integrate two datasets with PLS-DA

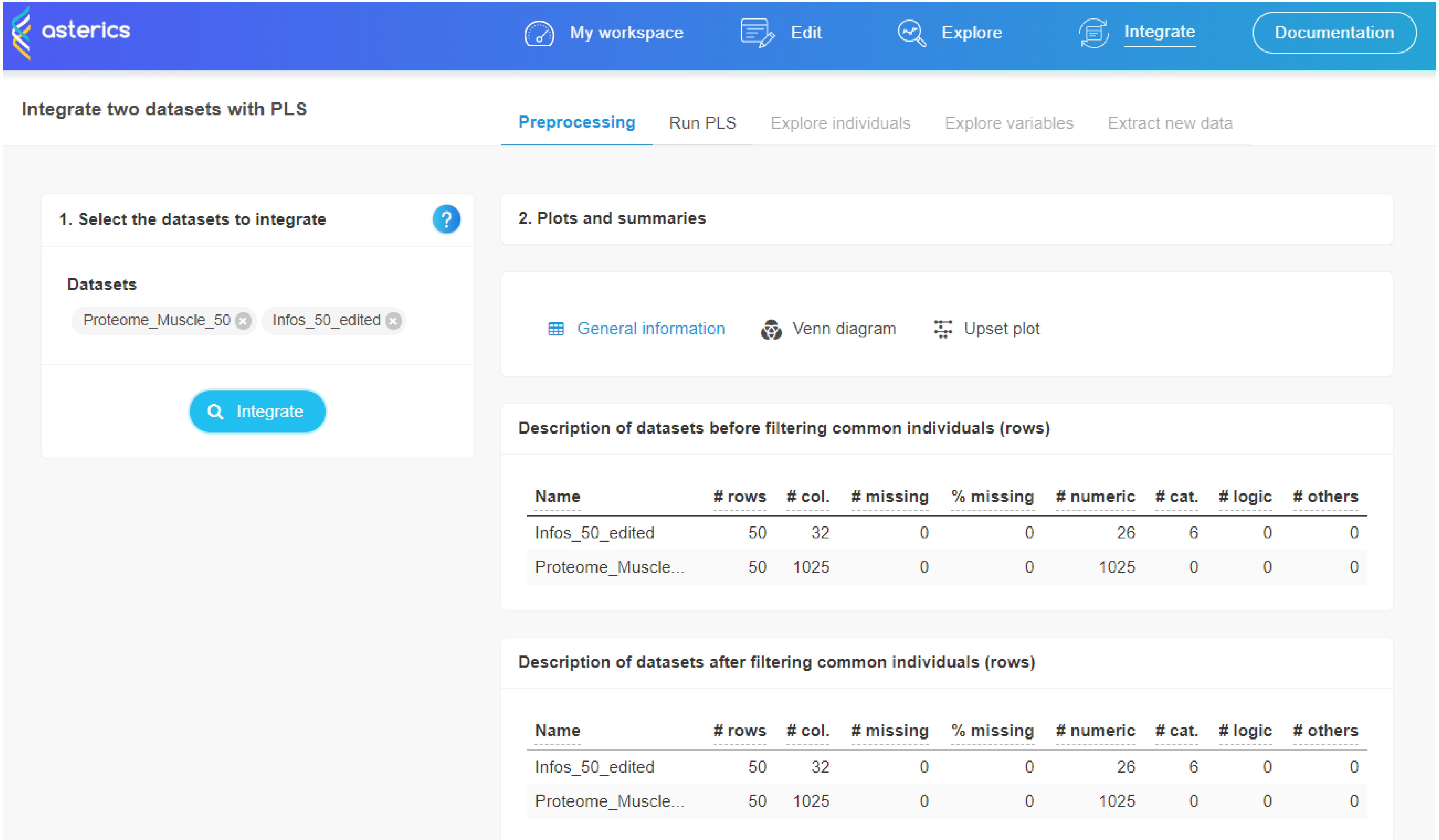
Performs Partial Least Squares analysis on two datasets.

Integrate datasets with MFA

Perform Multiple Factor Analysis on several datasets.

Differential analysis

Perform differential analysis for all numeric variables of a dataset.



Integrate two datasets with PLS

1. Select the datasets to integrate

Datasets

Proteome_Muscle_50 ⊗ Infos_50_edited ⊗

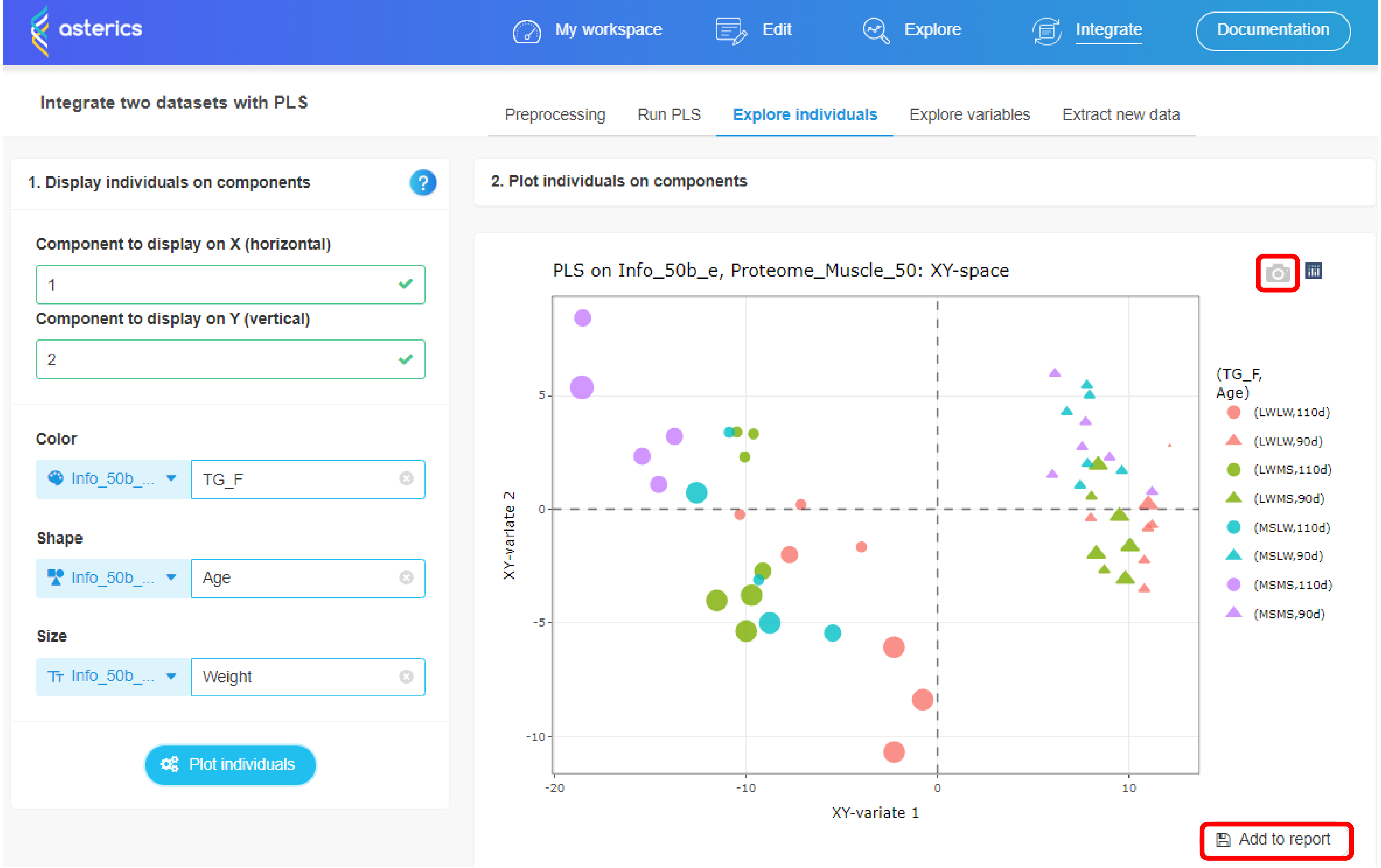
2. Plots and summaries

Description of datasets before filtering common individuals (rows)

Name	# rows	# col.	# missing	% missing	# numeric	# cat.	# logic	# others
Infos_50_edited	50	32	0	0	26	6	0	0
Proteome_Muscle...	50	1025	0	0	1025	0	0	0

Description of datasets after filtering common individuals (rows)

Name	# rows	# col.	# missing	% missing	# numeric	# cat.	# logic	# others
Infos_50_edited	50	32	0	0	26	6	0	0
Proteome_Muscle...	50	1025	0	0	1025	0	0	0



Integrate two datasets with PLS

1. Display individuals on components

Component to display on X (horizontal): 1

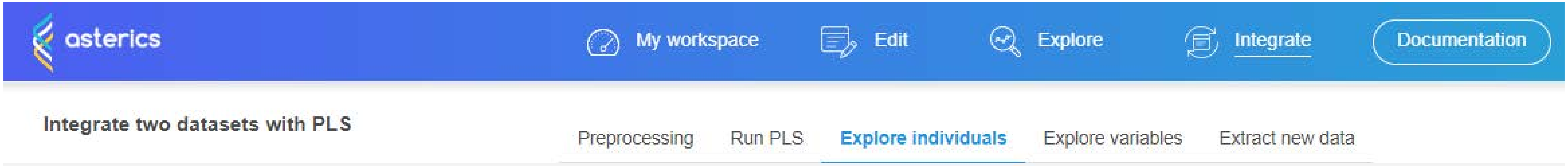
Component to display on Y (vertical): 2

Color: Info_50b_... TG_F

Shape: Info_50b_... Age

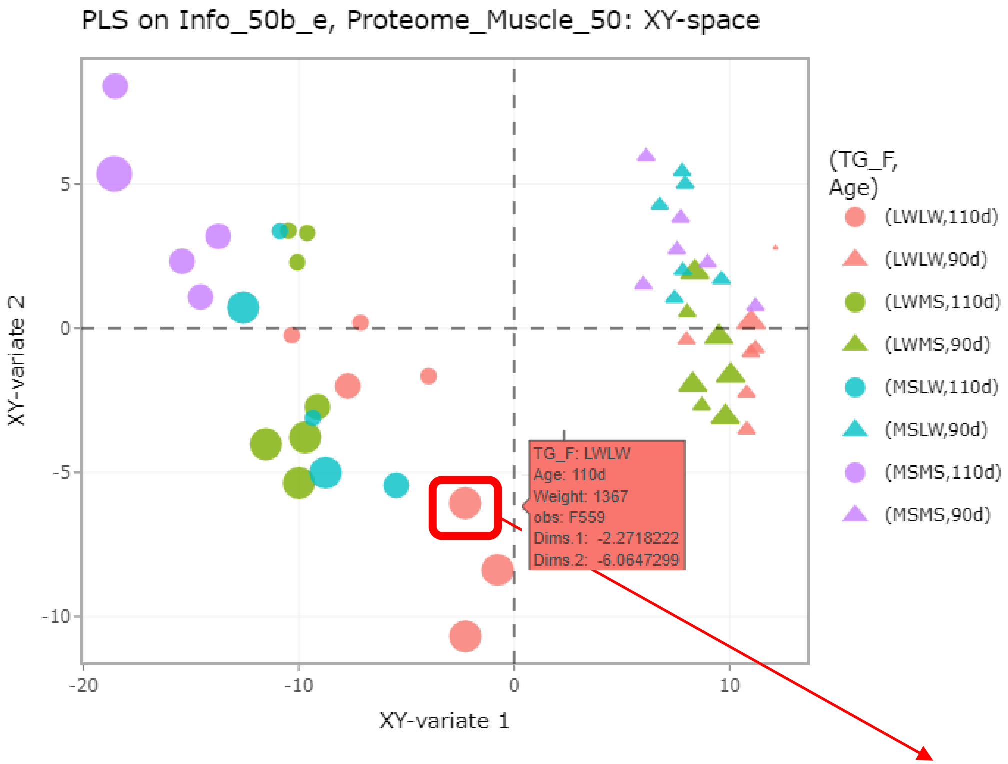
Size: Info_50b_... Weight

2. Plot individuals on components



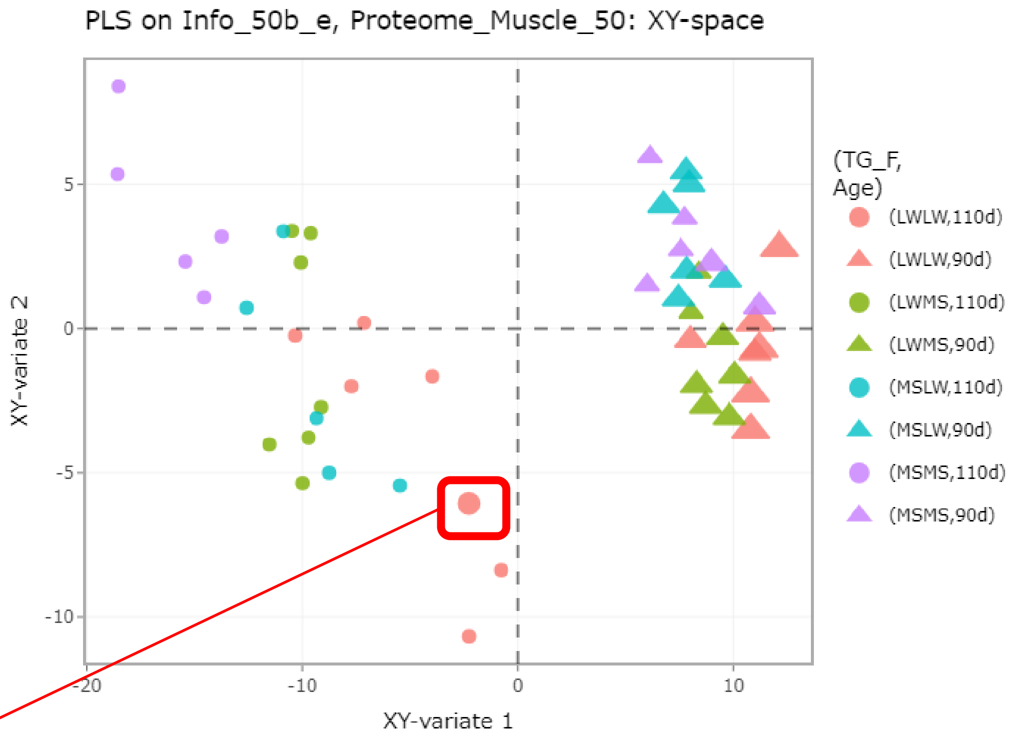
Integrate two datasets with PLS

Size = weight

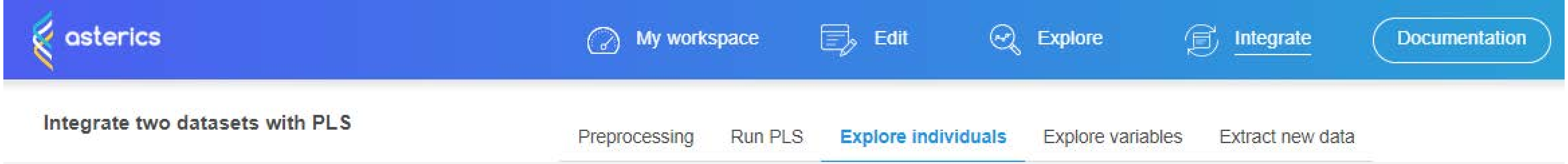


TG_F: LWLW
Age: 110d
Weight: 1367
obs: F559
Dims.1: -2.2718222
Dims.2: -6.0647299

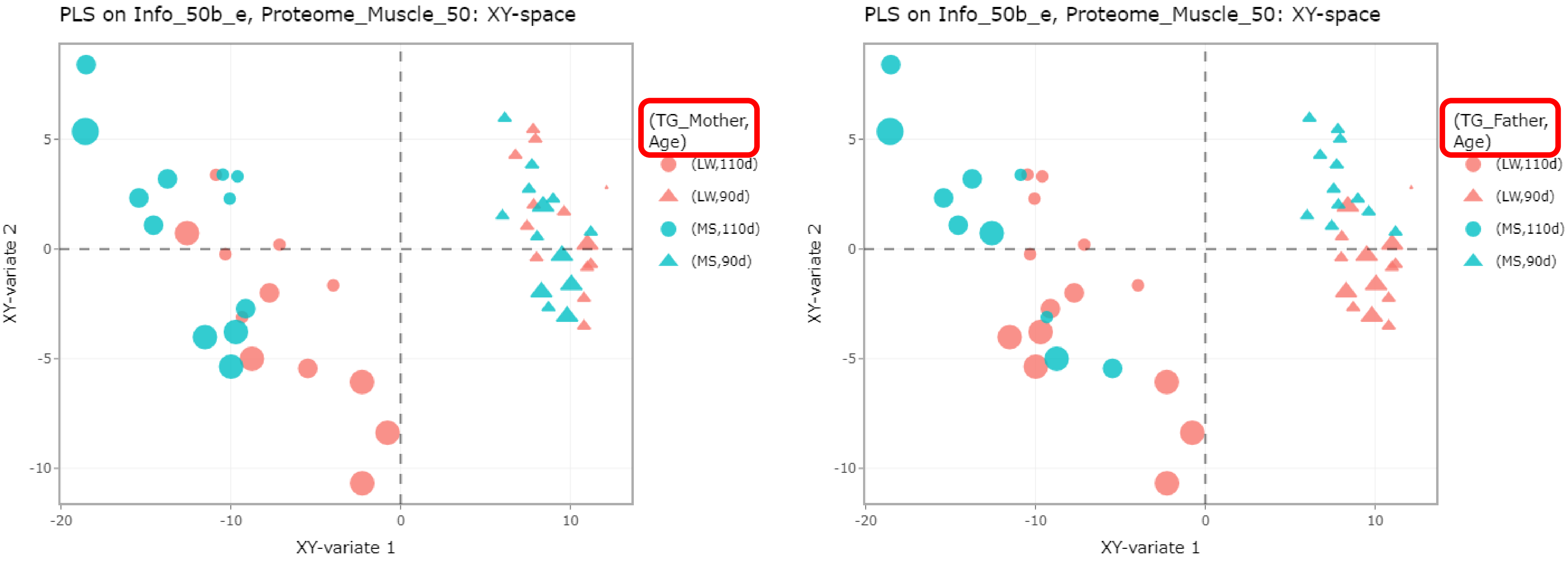
Size = MHC_E



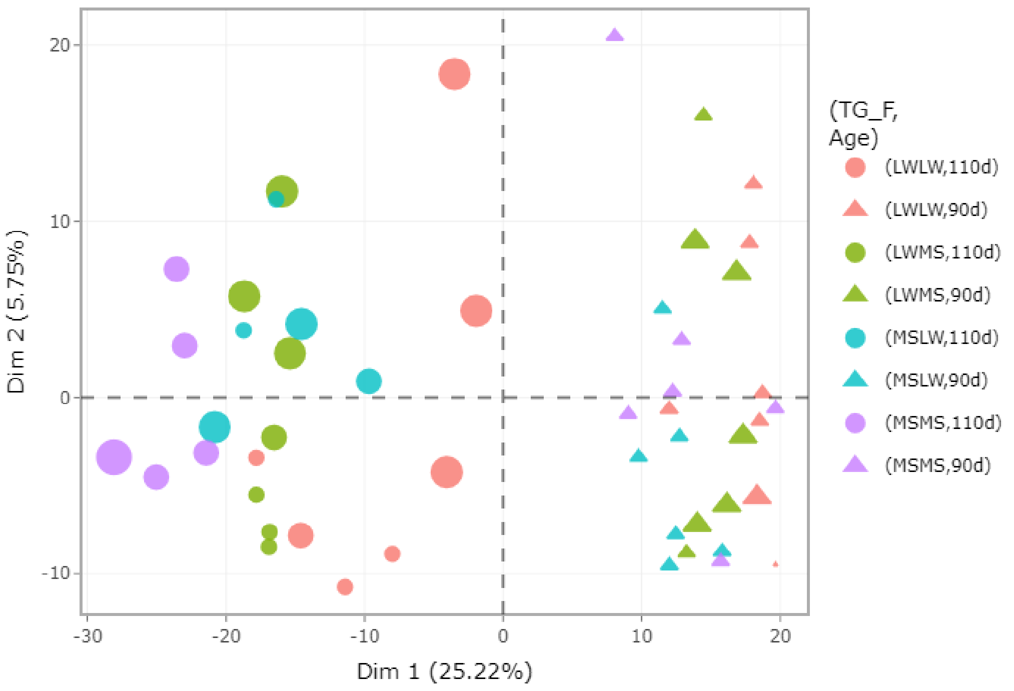
Heavy but developmental delay

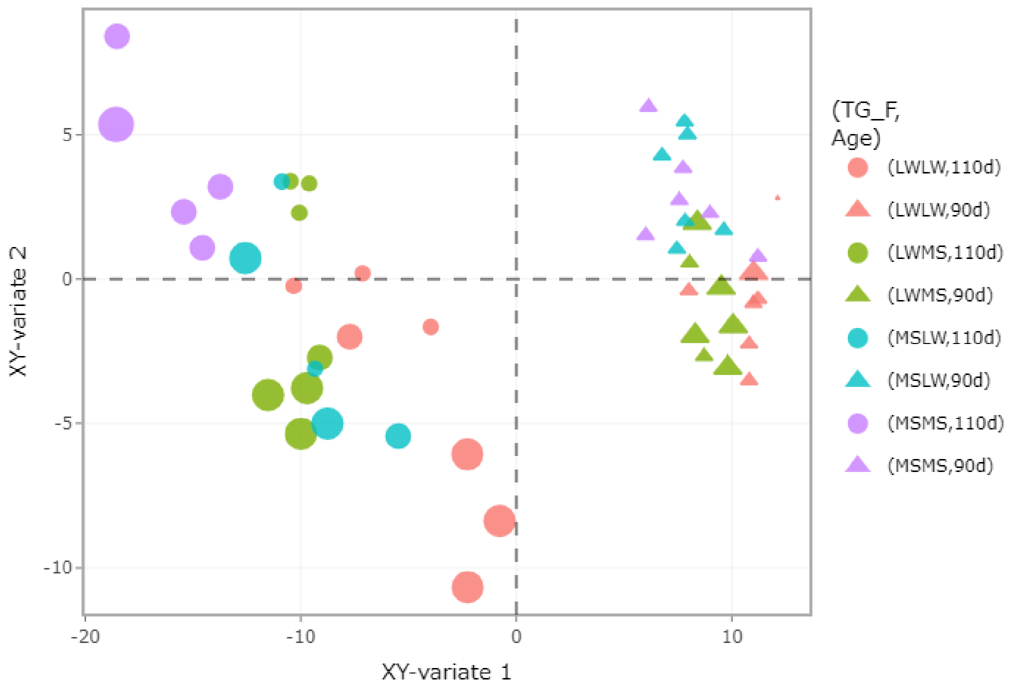

asterics
My workspace
Edit
Explore
Integrate
Documentation
Integrate two datasets with PLS
Preprocessing
Run PLS
Explore individuals
Explore variables
Extract new data

Size = weight

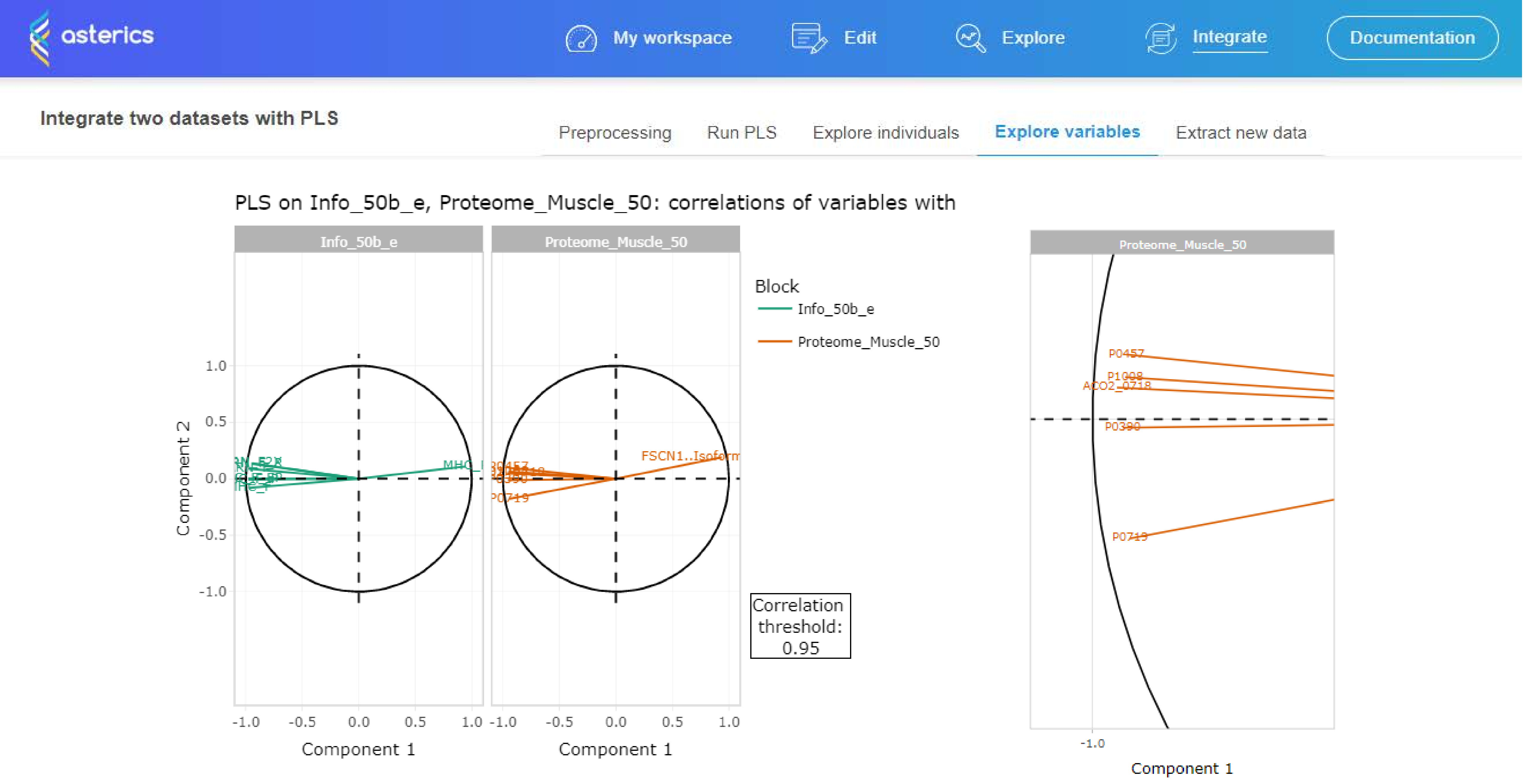

PLS on Info_50b_e, Proteome_Muscle_50: XY-space
XY-variate 2
XY-variate 1
5
0
-5
-10
-20
-10
0
10
(TG_Mother, Age)
(LW,110d)
(LW,90d)
(MS,110d)
(MS,90d)
PLS on Info_50b_e, Proteome_Muscle_50: XY-space
XY-variate 2
XY-variate 1
5
0
-5
-10
-20
-10
0
10
(TG_Father, Age)
(LW,110d)
(LW,90d)
(MS,110d)
(MS,90d)

PCA vs PLS


PCA on Proteome_Muscle_50: projection of individuals on the PC


PLS on Info_50b_e, Proteome_Muscle_50: XY-space

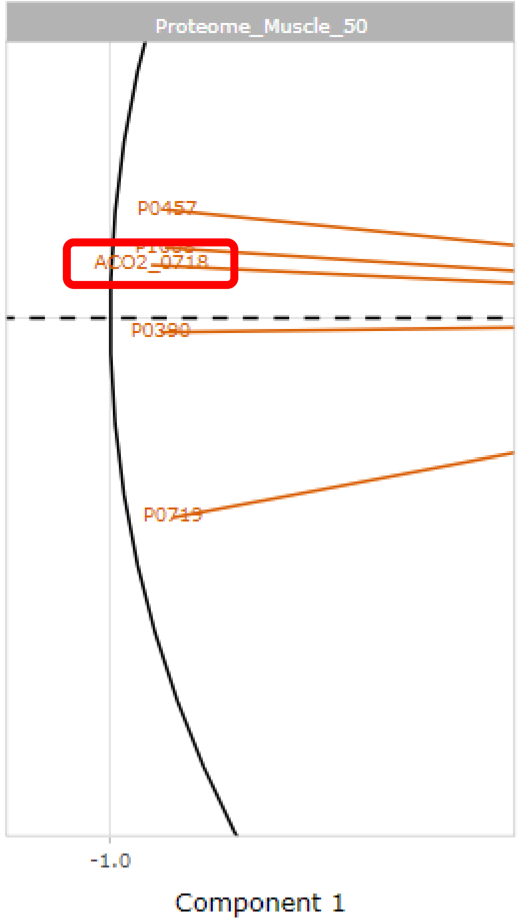
Size = weight



Integrate two datasets with PLS

PLS on Info_50b_e, Proteome_Muscle_50: correlations of variables with

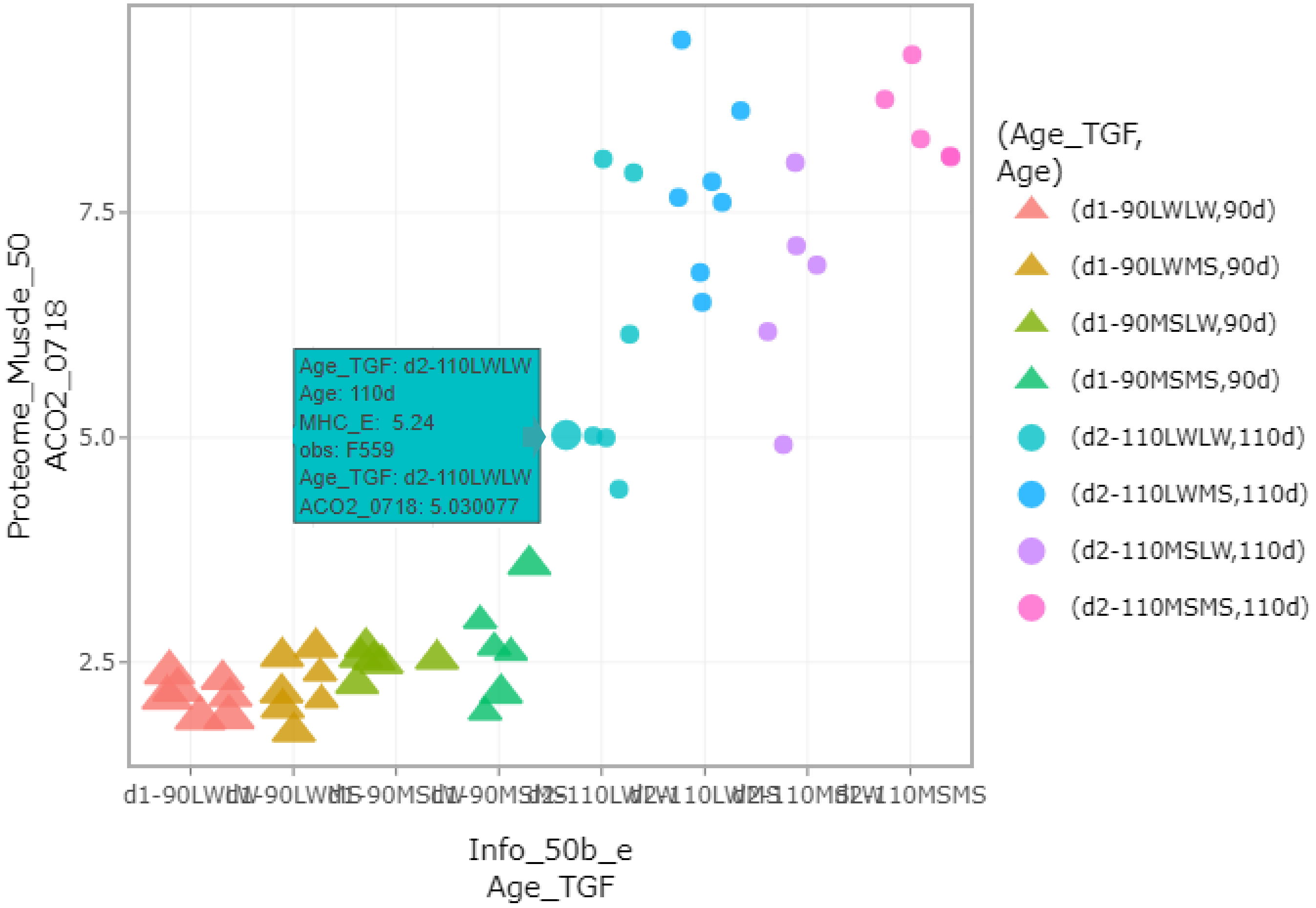
Correlation threshold: 0.95



asterics

Explore variables in a dataset

Size = MHC_E

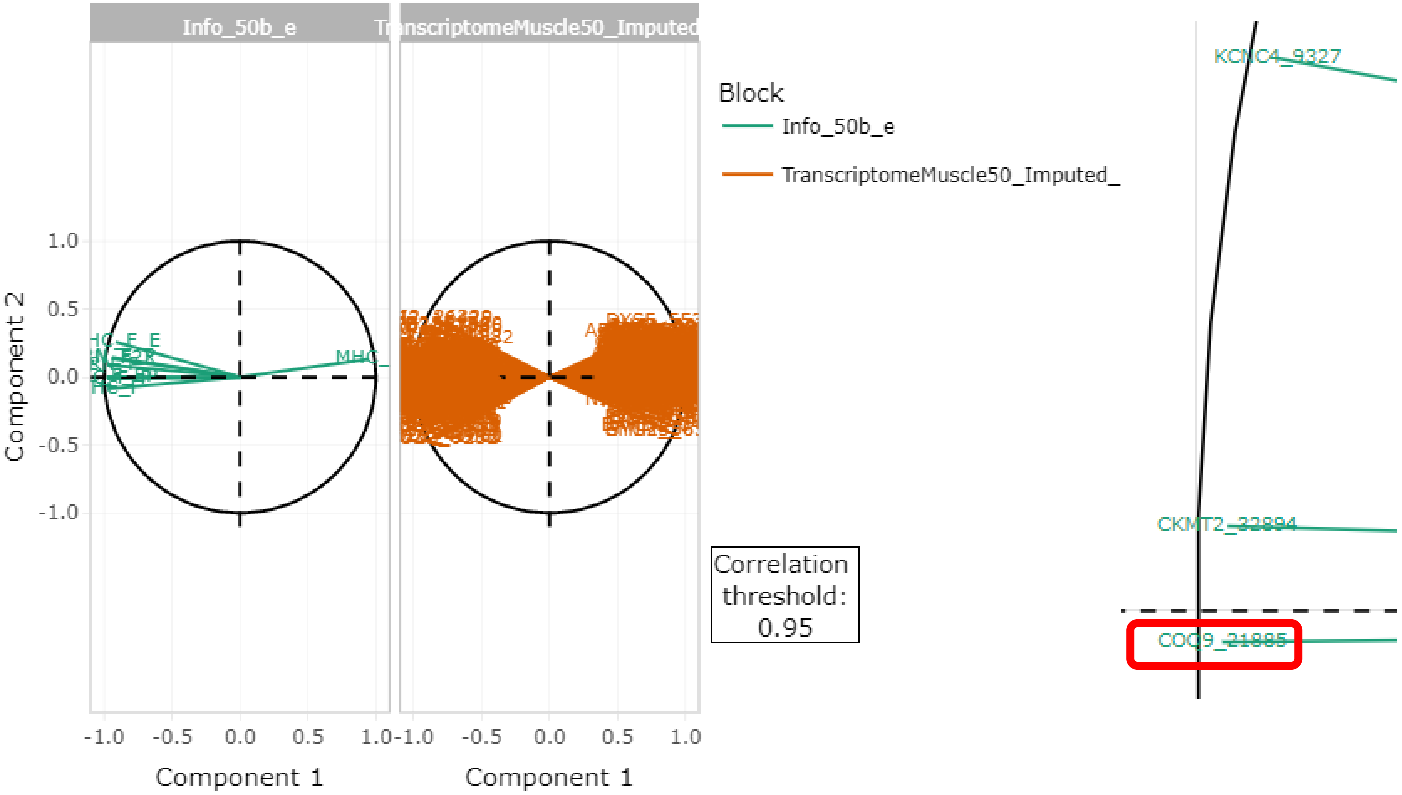


asterics My workspace Edit Explore Integrate Documentation

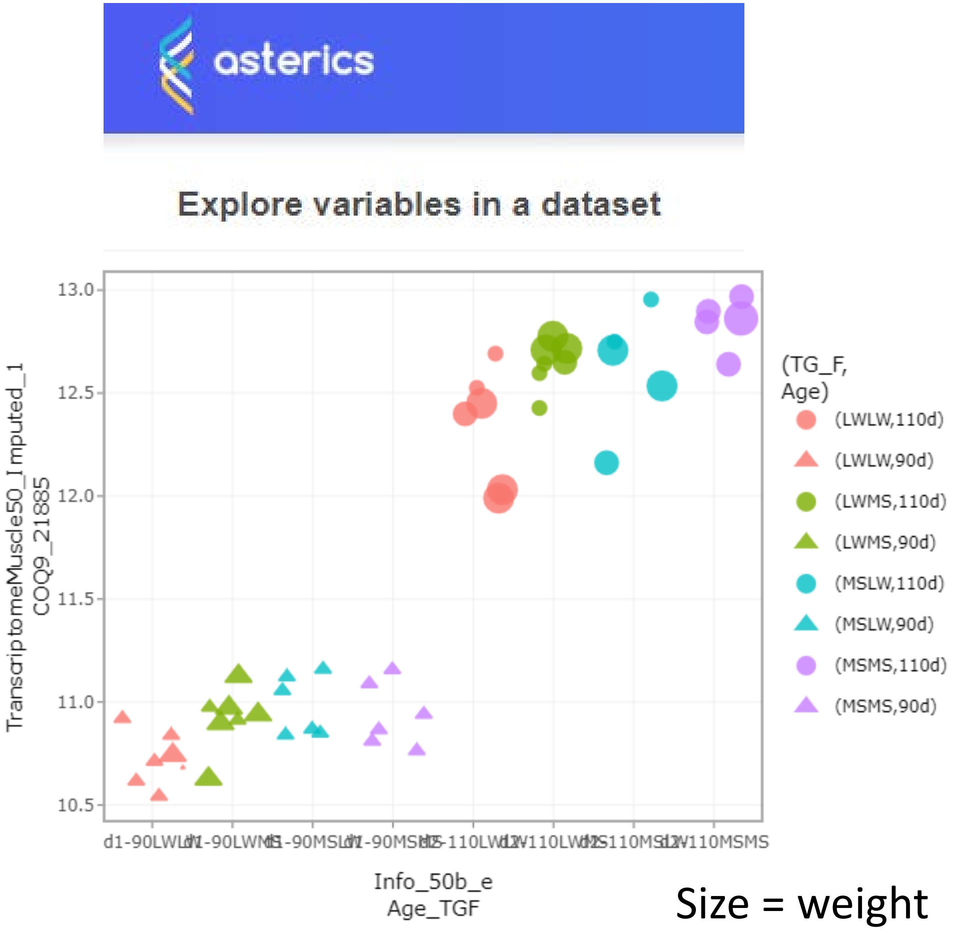
Integrate two datasets with PLS

Preprocessing Run PLS Explore individuals **Explore variables** Extract new data

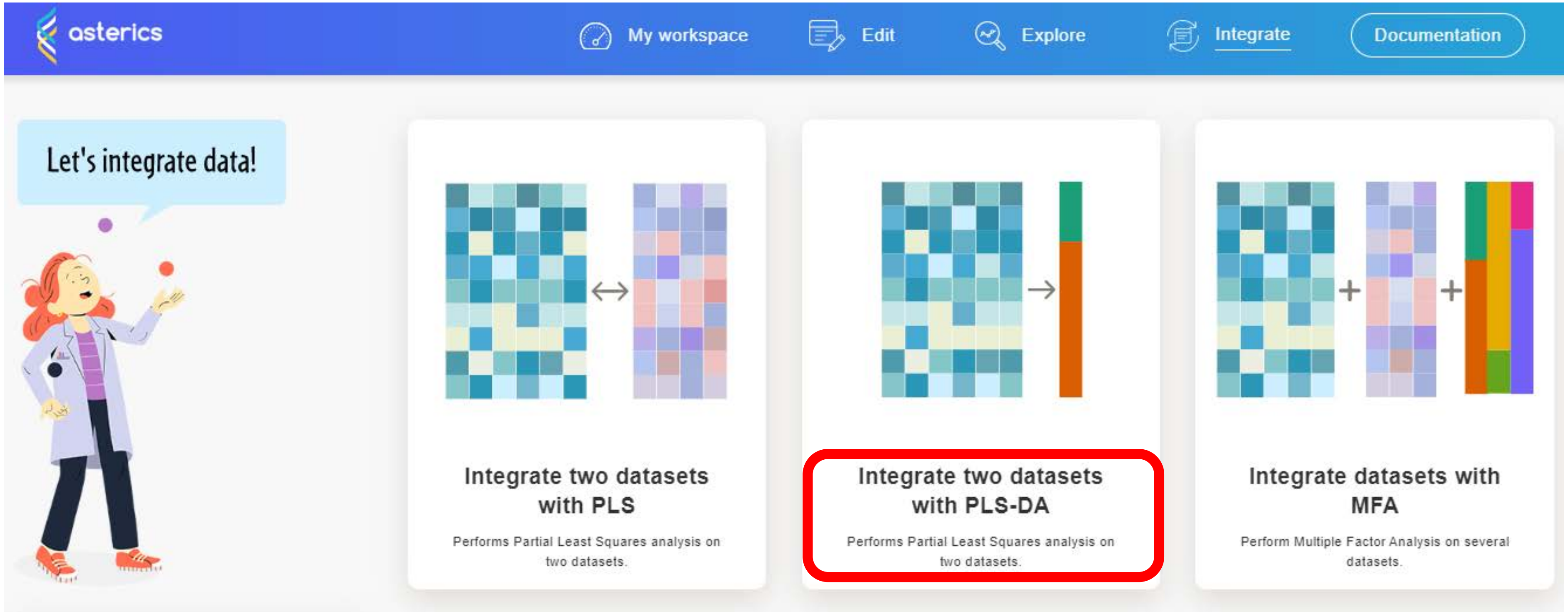
PLS on Info_50b_e, TranscriptomeMuscle50_Imputed_1: correlations of



Correlation threshold: 0.95



Size = weight



Integrate two datasets with PLS

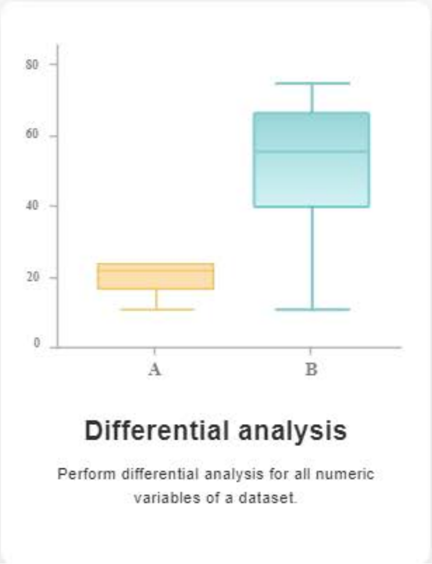
Performs Partial Least Squares analysis on two datasets.

Integrate two datasets with PLS-DA

Performs Partial Least Squares analysis on two datasets.

Integrate datasets with MFA

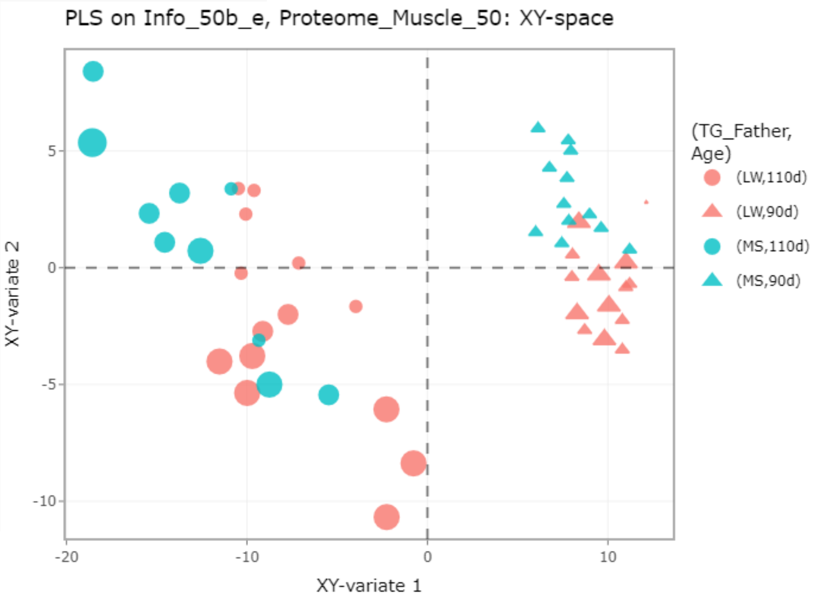
Perform Multiple Factor Analysis on several datasets.

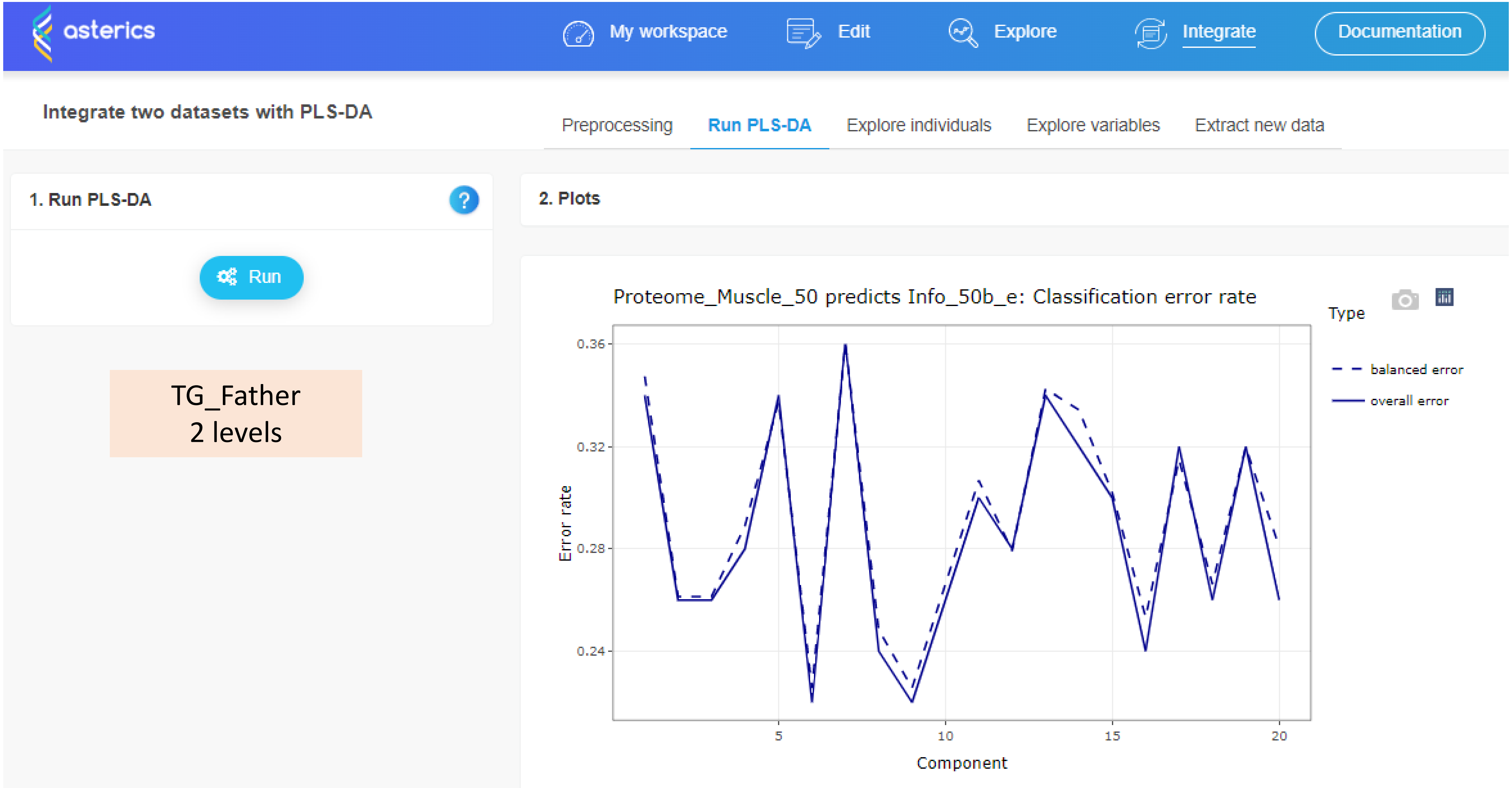


Differential analysis

Perform differential analysis for all numeric variables of a dataset.

It is easy to discriminate the age but what about the genotype?

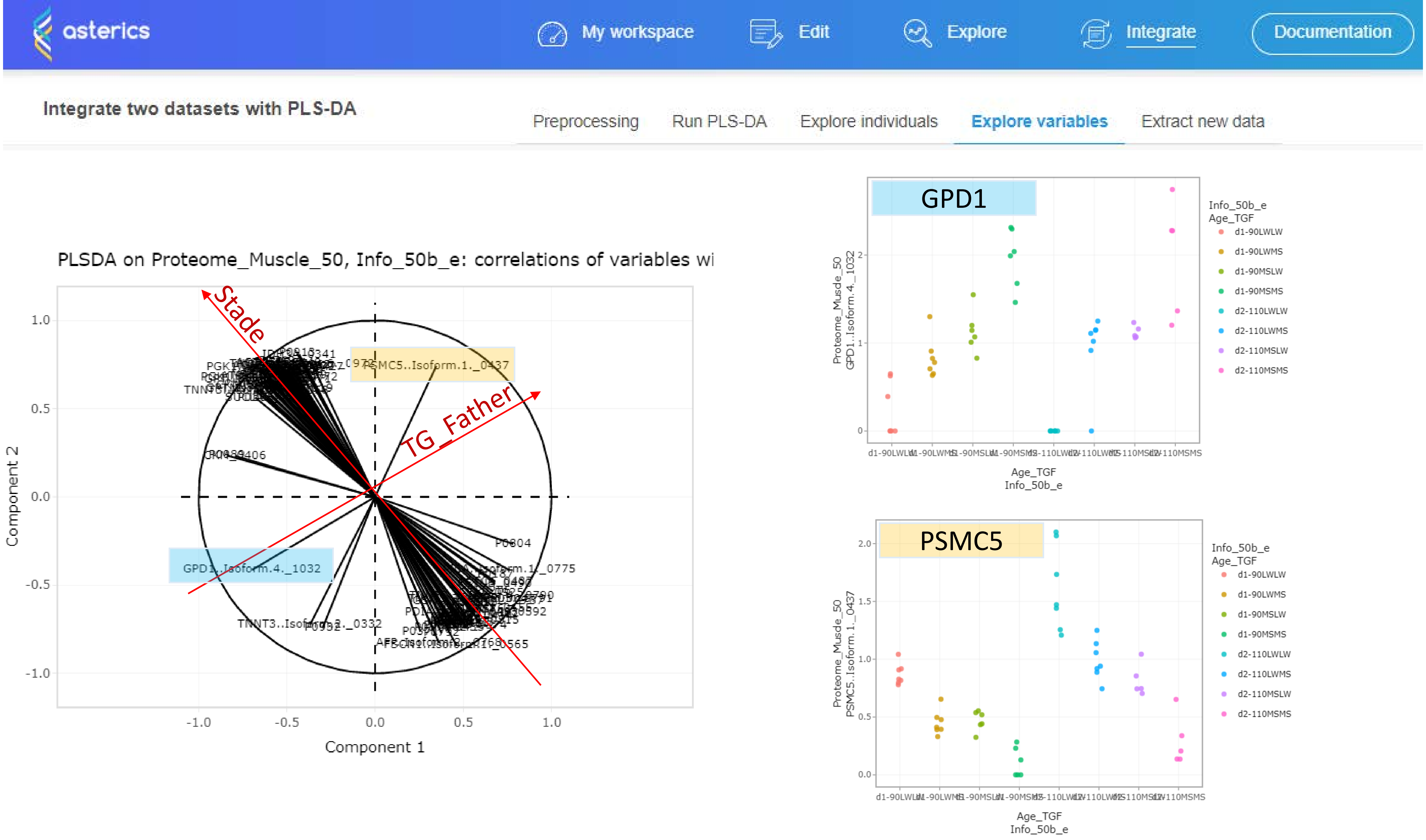


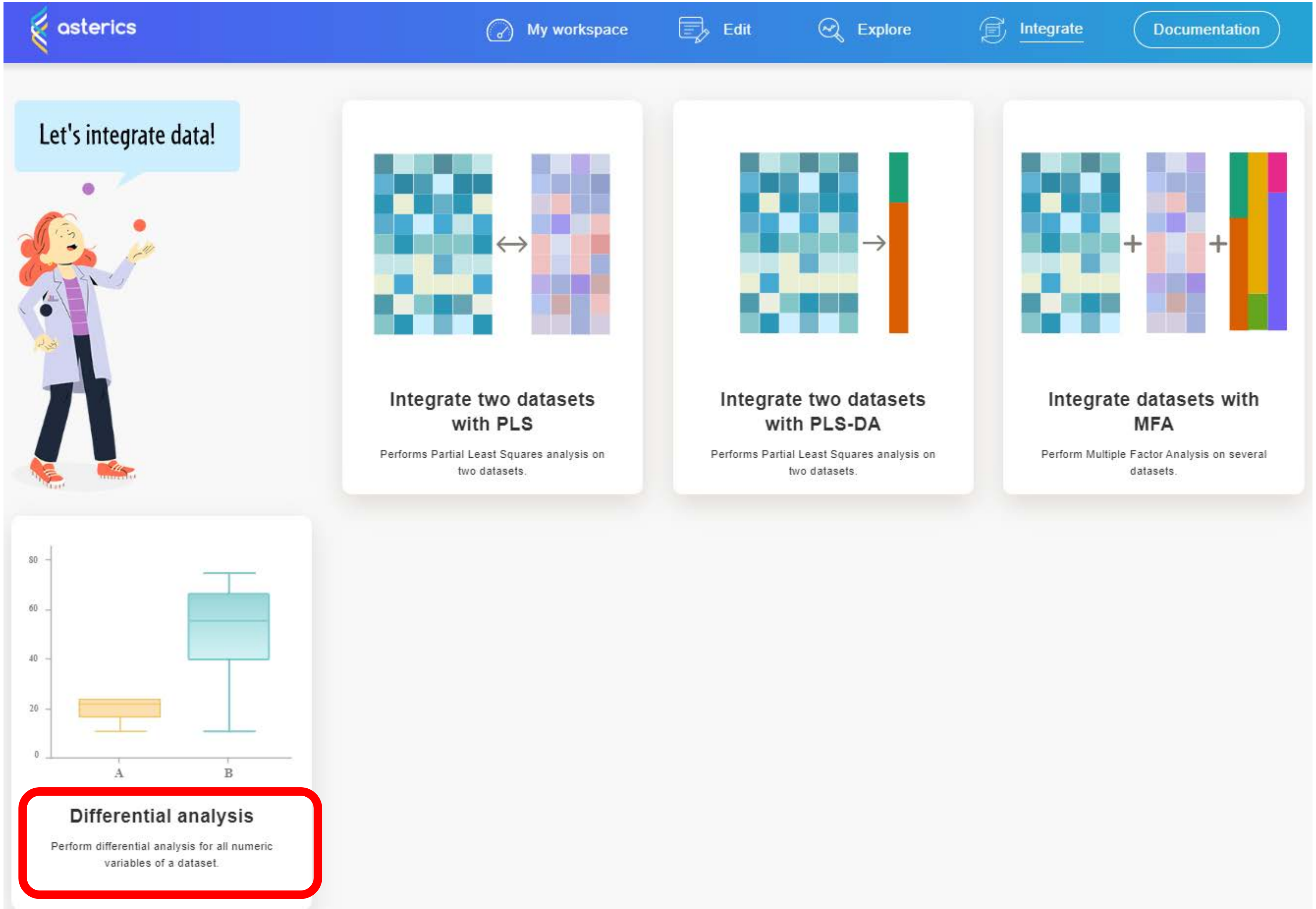


TG_Father
2 levels

Integrate two datasets with PLS-DA

Size = weight


asterics
My workspace
Edit
Explore
Integrate
Documentation
Integrate two datasets with PLS-DA
Preprocessing
Run PLS-DA
Explore individuals
Explore variables
Extract new data
PLSDA on Proteome_Muscle_50, Info_50b_e: correlations of variables wi
Stade
TG_Father
PSMC5..Isoform.1._0437
GPD1..Isoform.4._1032
TNNT3..Isoform.2._0332
P0804
P0932
P0406
Component 2
Component 1
1.0
0.5
0.0
-0.5
-1.0
GPD1
Proteome_Muscle_50
GPD1..Isoform.4._1032
Age_TGF
Info_50b_e
d1-90LWLW
d1-90LWMS
d1-90MSLW
d1-90MSMS
d2-110LWLW
d2-110LWMS
d2-110MSLW
d2-110MSMS
0
1
2
PSMC5
Proteome_Muscle_50
PSMC5..Isoform.1._0437
0.0
0.5
1.0
1.5
2.0



Integrate two datasets with PLS

Performs Partial Least Squares analysis on two datasets.

Integrate two datasets with PLS-DA

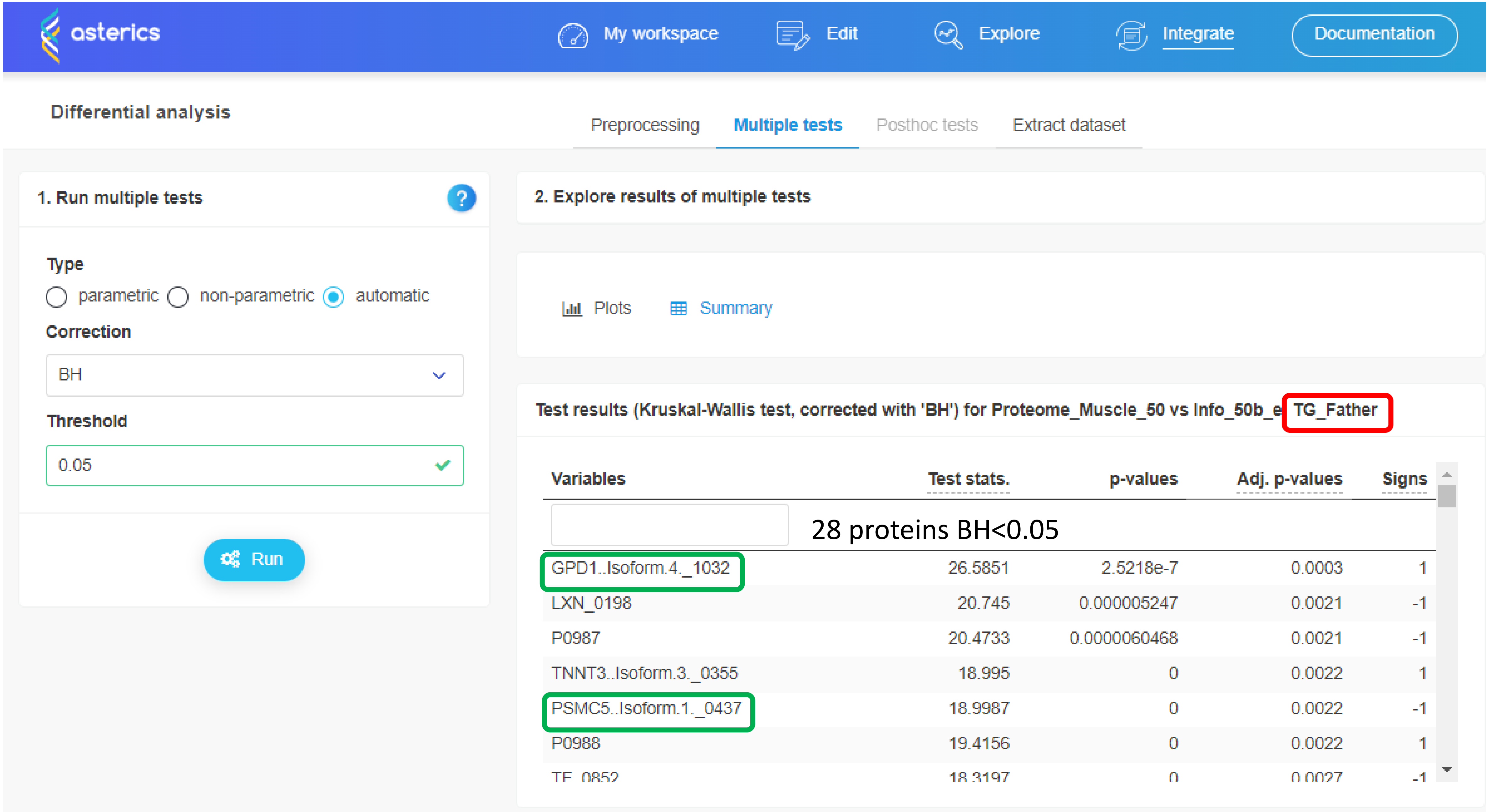
Performs Partial Least Squares analysis on two datasets.

Integrate datasets with MFA

Perform Multiple Factor Analysis on several datasets.

Differential analysis

Perform differential analysis for all numeric variables of a dataset.



Differential analysis

1. Run multiple tests

Type

○ parametric ○ non-parametric ● automatic

Correction

BH

Threshold

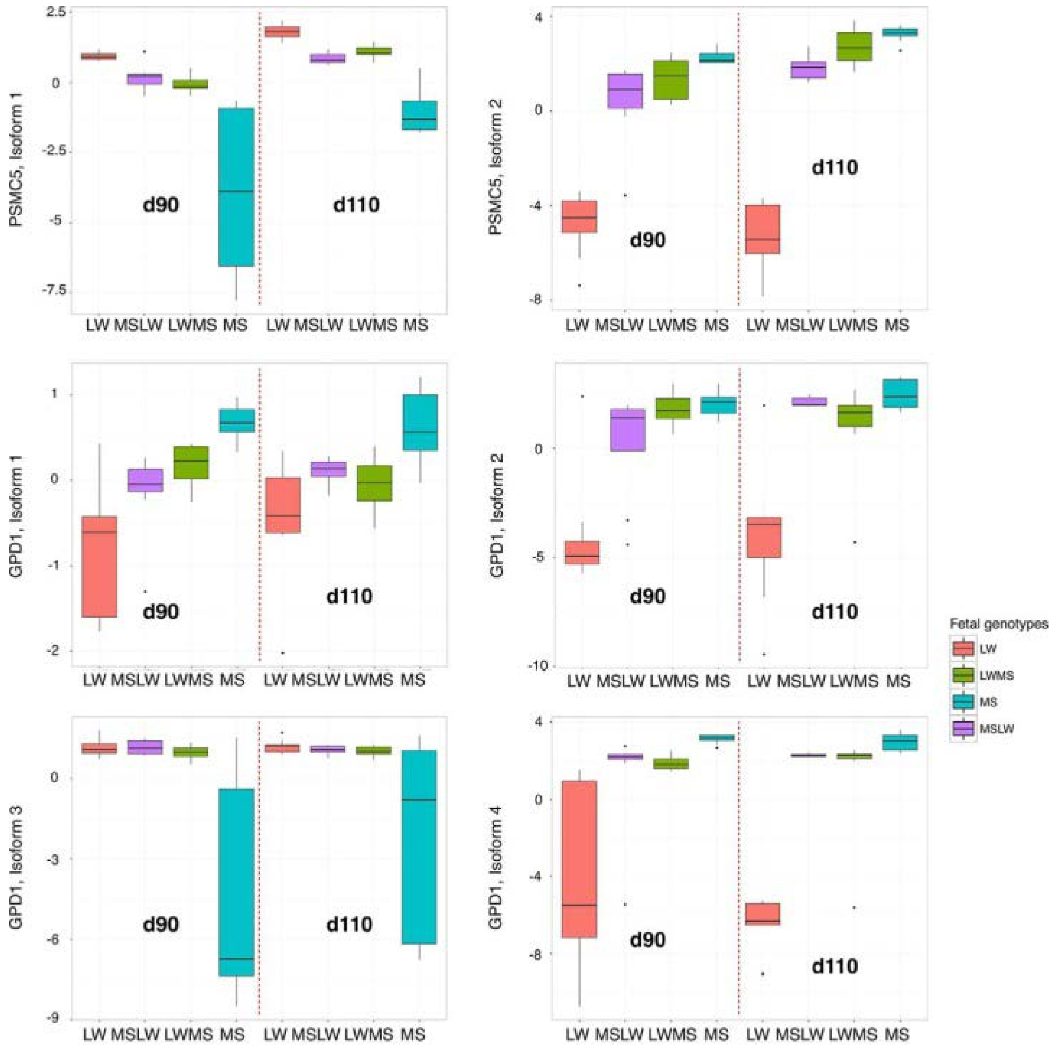
0.05

2. Explore results of multiple tests

Test results (Kruskal-Wallis test, corrected with 'BH') for Proteome_Muscle_50 vs Info_50b_e TG_Father

28 proteins BH<0.05

Variables	Test stats.	p-values	Adj. p-values	Signs
GPD1..Isoform.4._1032	26.5851	2.5218e-7	0.0003	1
LXN_0198	20.745	0.000005247	0.0021	-1
P0987	20.4733	0.0000060468	0.0021	-1
TNNT3..Isoform.3._0355	18.995	0	0.0022	1
PSMC5..Isoform.1._0437	18.9987	0	0.0022	-1
P0988	19.4156	0	0.0022	1
TF_0852	18.3197	0	0.0027	-1



PSMC5,
Proteasome 26S Regulatory Subunit, ATP-dependent degradation of ubiquitinated proteins, 5
2 isoforms

PGD1
Glycerol-3-phosphate dehydrogenase 1, cytoplasmic
4 isoforms

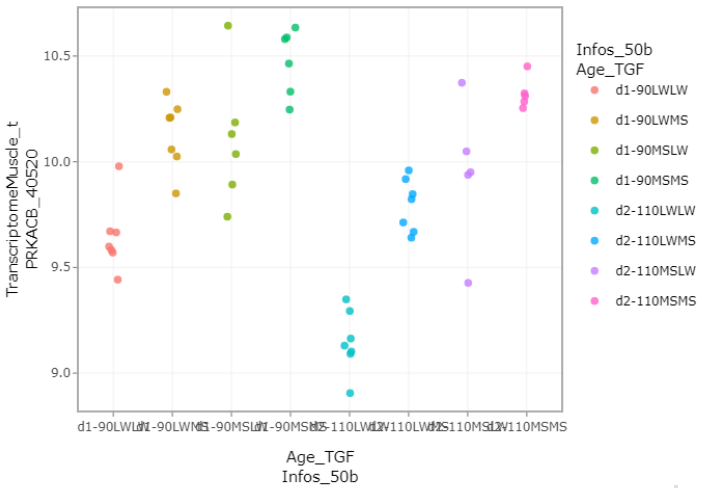
Voillet et al., 2018

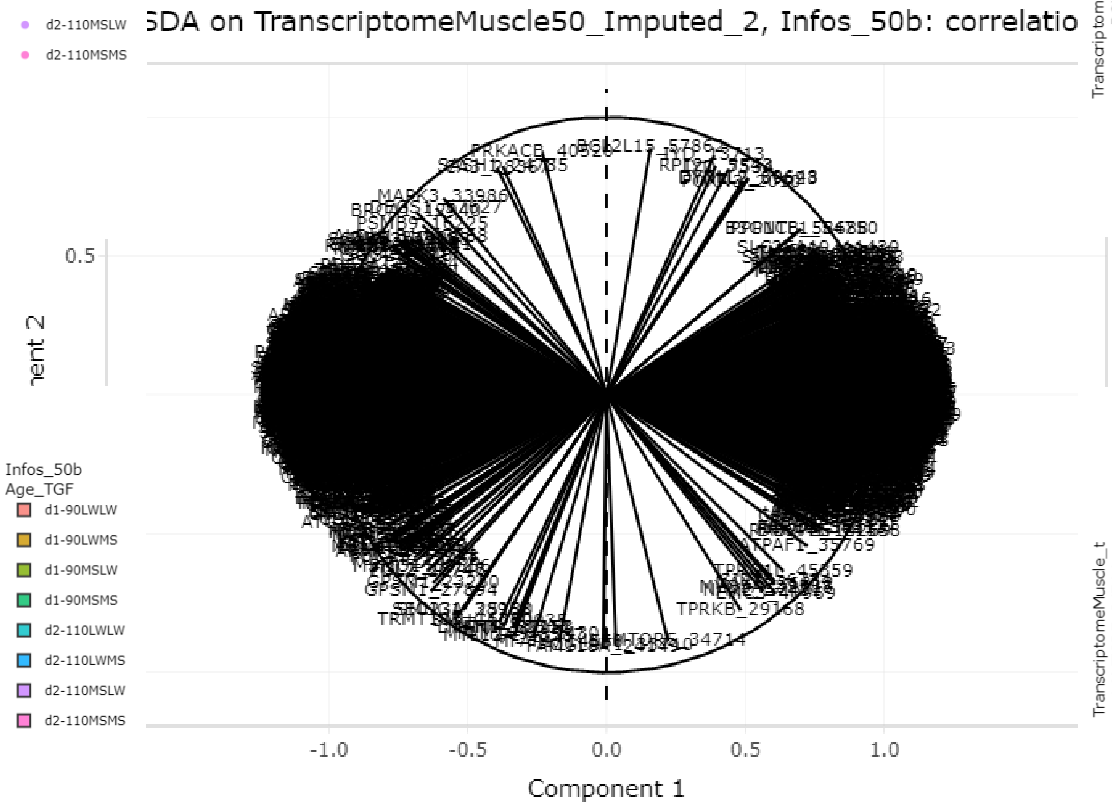
Integrate two datasets with PLS-DA

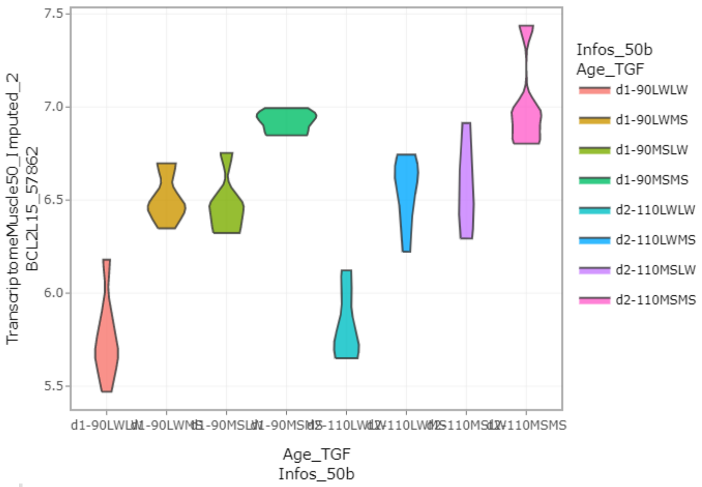
Age_TGF
8 levels

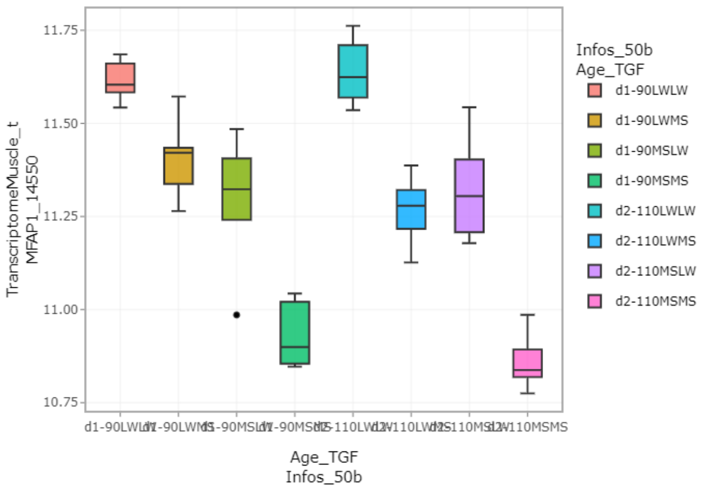
TranscriptomeMuscle50_Imputed_2 predicts Infos_50b: Classification err

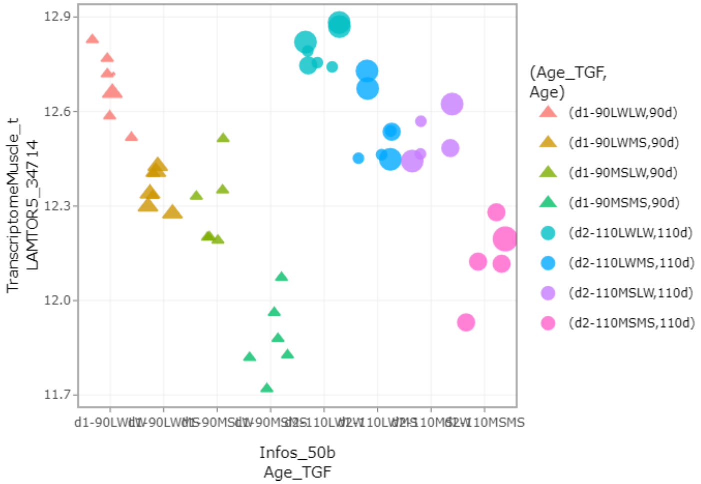
PLSDA on TranscriptomeMuscle50_Imputed_2, Infos_50b: X-space

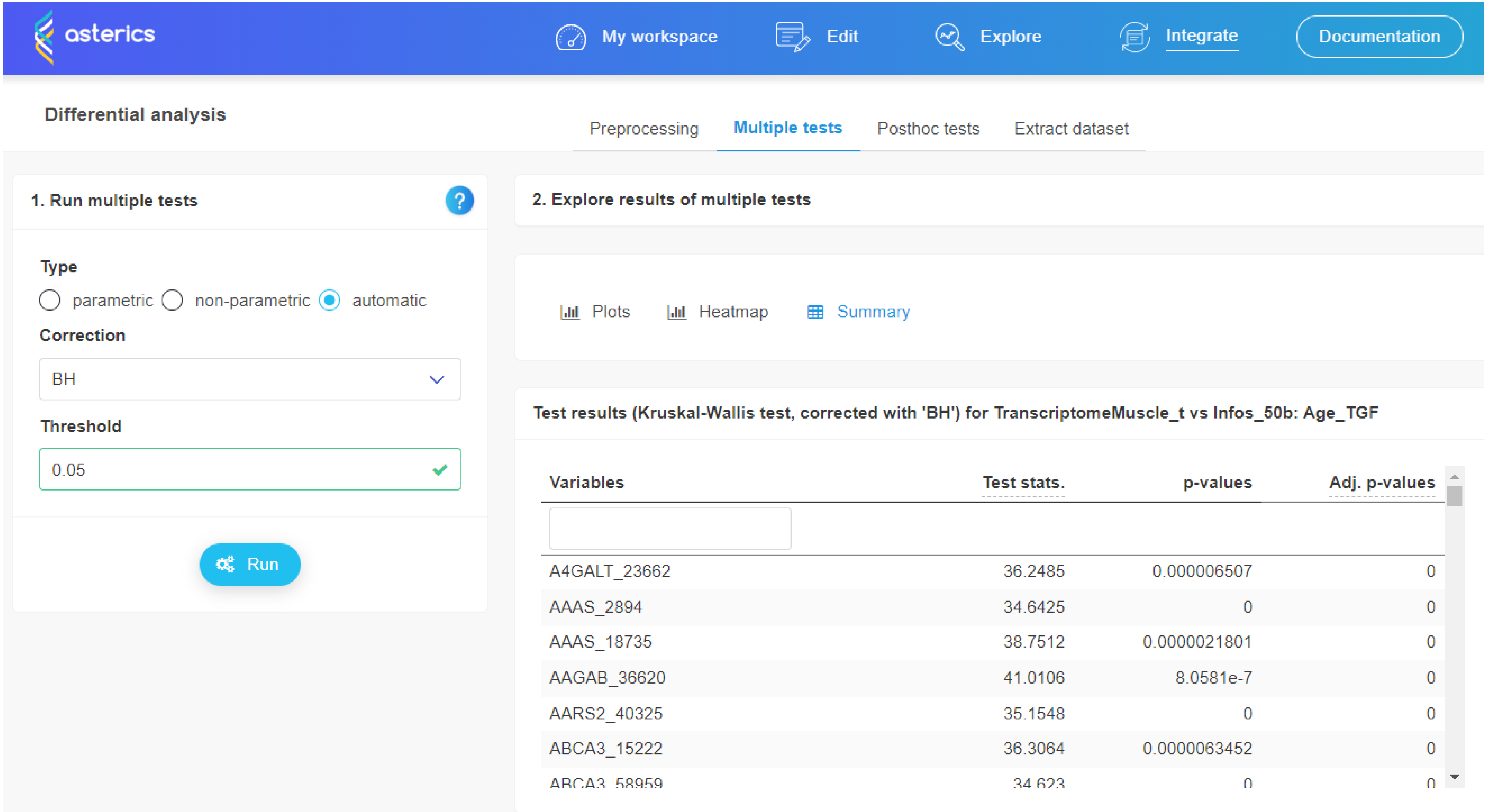

TranscriptomeMuscle_t
PRKACB_40520
10.5
10.0
9.5
9.0
d1-90LWLW
d1-90LWMS
d1-90MSLW
d1-90MSMS
d2-110LWLW
d2-110LWMS
d2-110MSLW
d2-110MSMS
Age_TGF
Infos_50b
Infos_50b
Age_TGF
d1-90LWLW
d1-90LWMS
d1-90MSLW
d1-90MSMS
d2-110LWLW
d2-110LWMS
d2-110MSLW
d2-110MSMS


SDA on TranscriptomeMuscle50_Imputed_2, Infos_50b: correlatio
PRKACB_40520
BCL2L15_57862
MAPK3_33986
ATPAF1_35769
TPRKB_29168
LAMTOR5_34714
0.5
ent 2
-1.0
-0.5
0.0
0.5
1.0
Component 1


TranscriptomeMuscle50_Imputed_2
BCL2L15_57862
7.5
7.0
6.5
6.0
5.5
d1-90LWLW
d1-90LWMS
d1-90MSLW
d1-90MSMS
d2-110LWLW
d2-110LWMS
d2-110MSLW
d2-110MSMS
Age_TGF
Infos_50b
Infos_50b
Age_TGF
d1-90LWLW
d1-90LWMS
d1-90MSLW
d1-90MSMS
d2-110LWLW
d2-110LWMS
d2-110MSLW
d2-110MSMS


TranscriptomeMuscle_t
MFAP1_14550
11.75
11.50
11.25
11.00
10.75
d1-90LWLW
d1-90LWMS
d1-90MSLW
d1-90MSMS
d2-110LWLW
d2-110LWMS
d2-110MSLW
d2-110MSMS
Age_TGF
Infos_50b
Infos_50b
Age_TGF
d1-90LWLW
d1-90LWMS
d1-90MSLW
d1-90MSMS
d2-110LWLW
d2-110LWMS
d2-110MSLW
d2-110MSMS


TranscriptomeMuscle_t
LAMTOR5_34714
12.9
12.6
12.3
12.0
11.7
d1-90LWLW
d1-90LWMS
d1-90MSLW
d1-90MSMS
d2-110LWLW
d2-110LWMS
d2-110MSLW
d2-110MSMS
Infos_50b
Age_TGF
(Age_TGF,
Age)
(d1-90LWLW,90d)
(d1-90LWMS,90d)
(d1-90MSLW,90d)
(d1-90MSMS,90d)
(d2-110LWLW,110d)
(d2-110LWMS,110d)
(d2-110MSLW,110d)
(d2-110MSMS,110d)



Differential analysis

1. Run multiple tests

Type

○ parametric ○ non-parametric ● automatic

Correction

BH

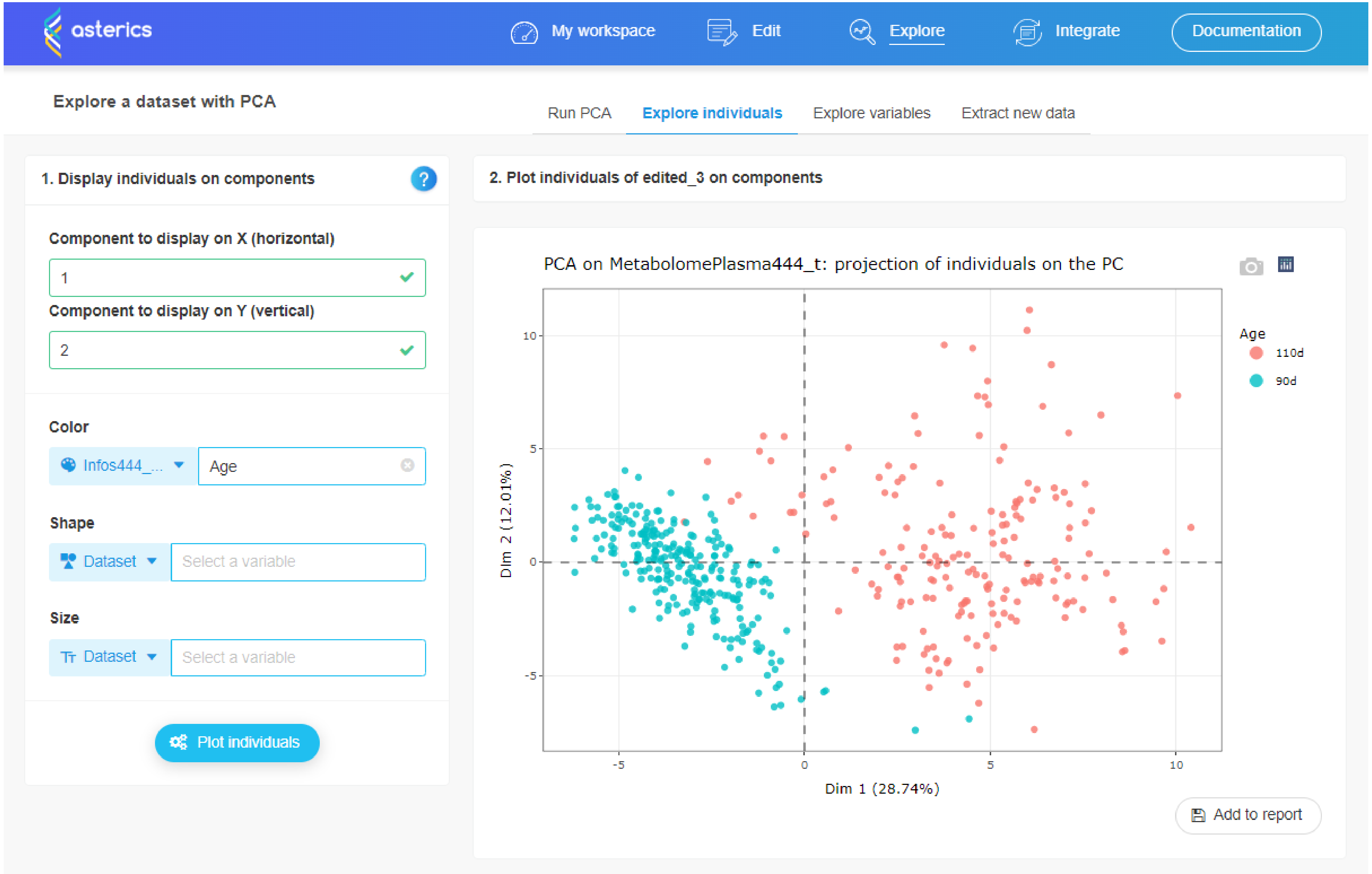
Threshold

0.05

2. Explore results of multiple tests

Test results (Kruskal-Wallis test, corrected with 'BH') for TranscriptomeMuscle_t vs Infos_50b: Age_TGF

Variables	Test stats.	p-values	Adj. p-values
A4GALT_23662	36.2485	0.000006507	0
AAAS_2894	34.6425	0	0
AAAS_18735	38.7512	0.0000021801	0
AAGAB_36620	41.0106	8.0581e-7	0
AARS2_40325	35.1548	0	0
ABCA3_15222	36.3064	0.0000063452	0
ABCA3_58959	34.623	0	0



Explore a dataset with PCA

1. Display individuals on components

Component to display on X (horizontal): 1

Component to display on Y (vertical): 2

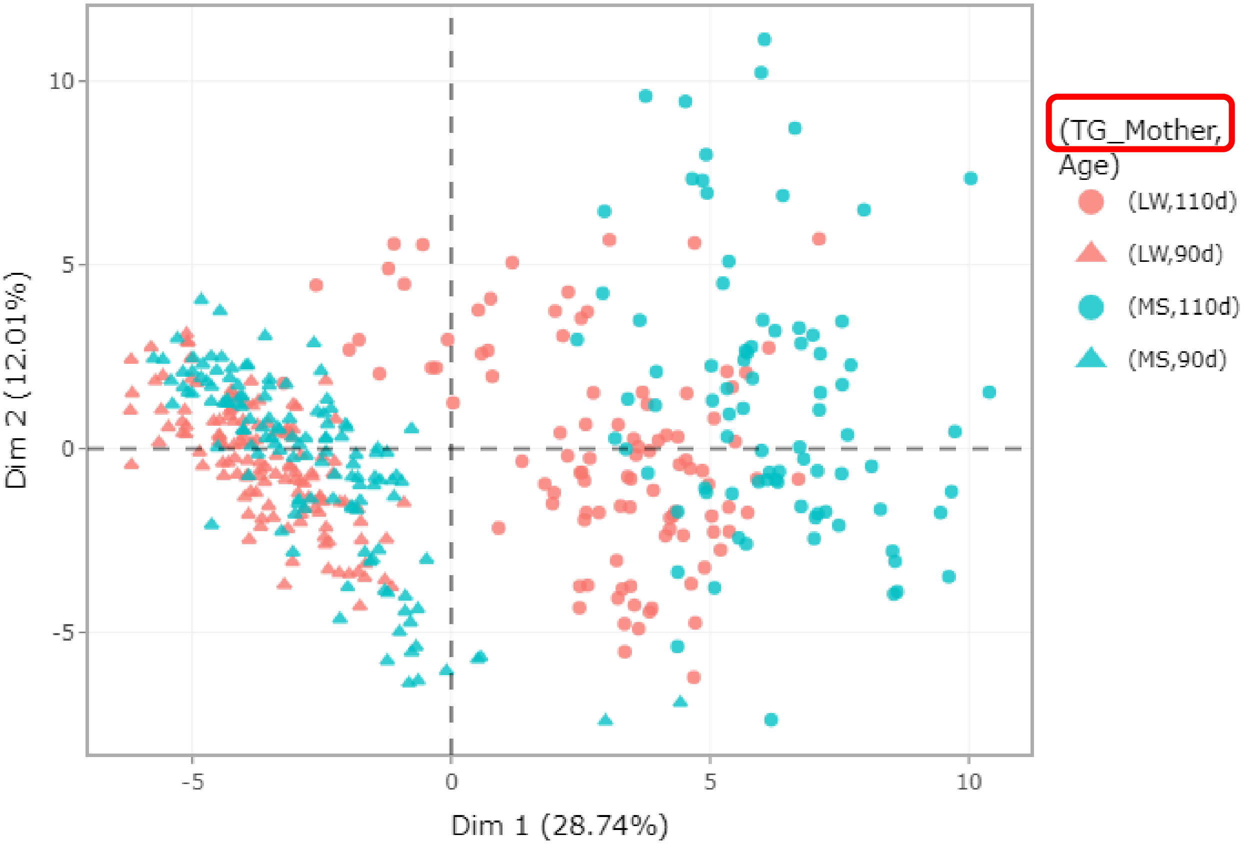
Color: Infos444_... Age

Shape: Dataset Select a variable

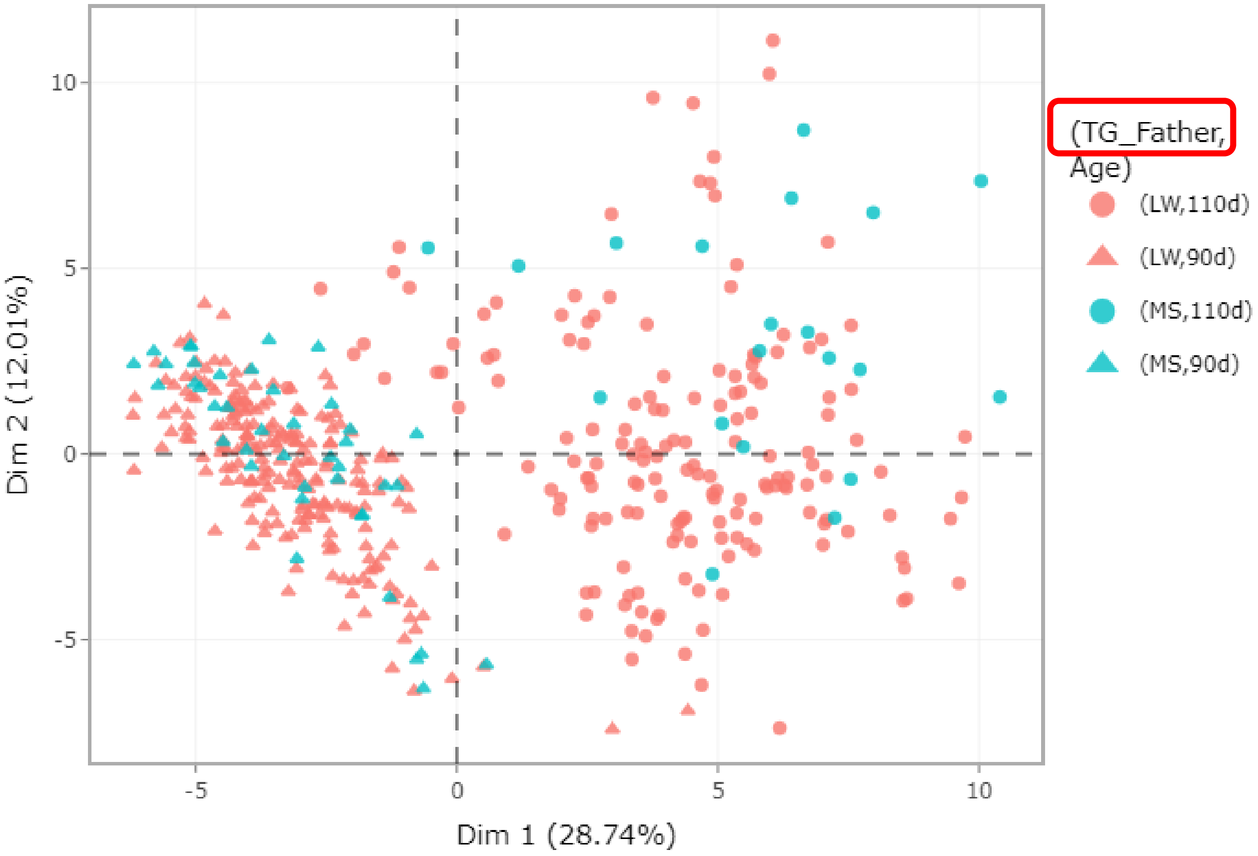
Size: Dataset Select a variable

2. Plot individuals of edited_3 on components

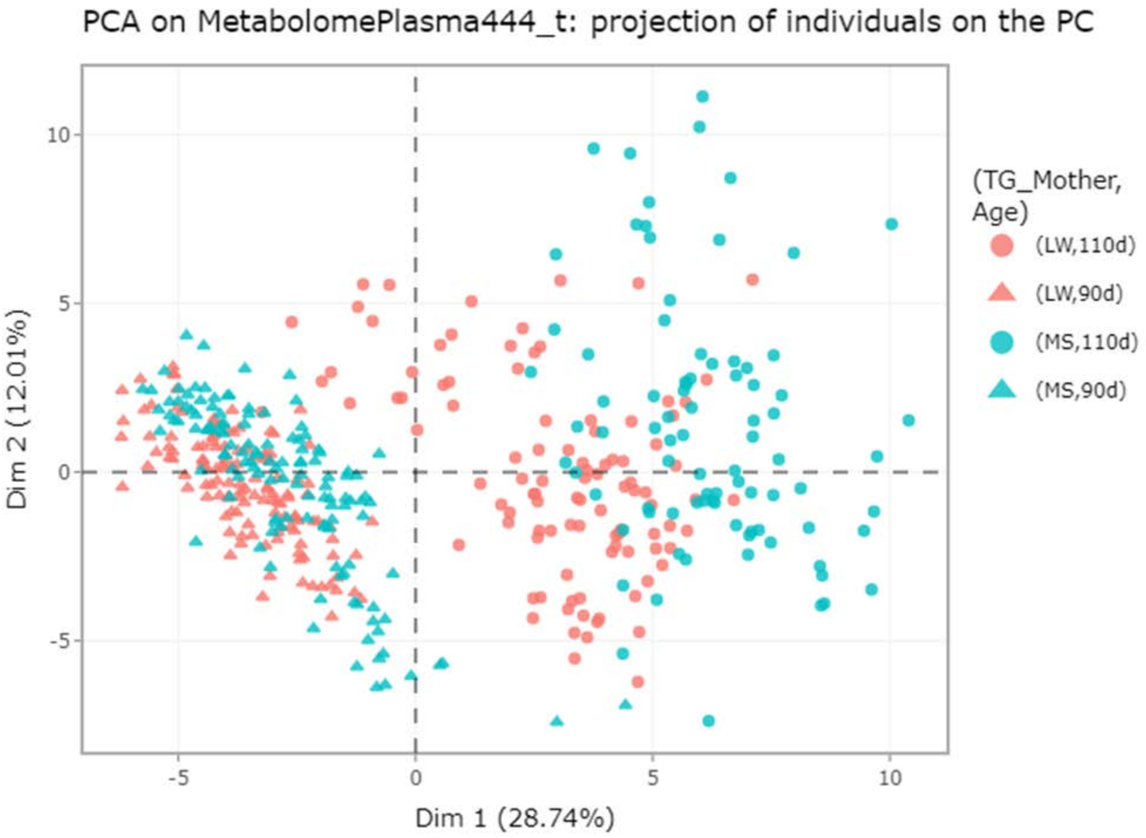
PCA on MetabolomePlasma444_t: projection of individuals on the PC

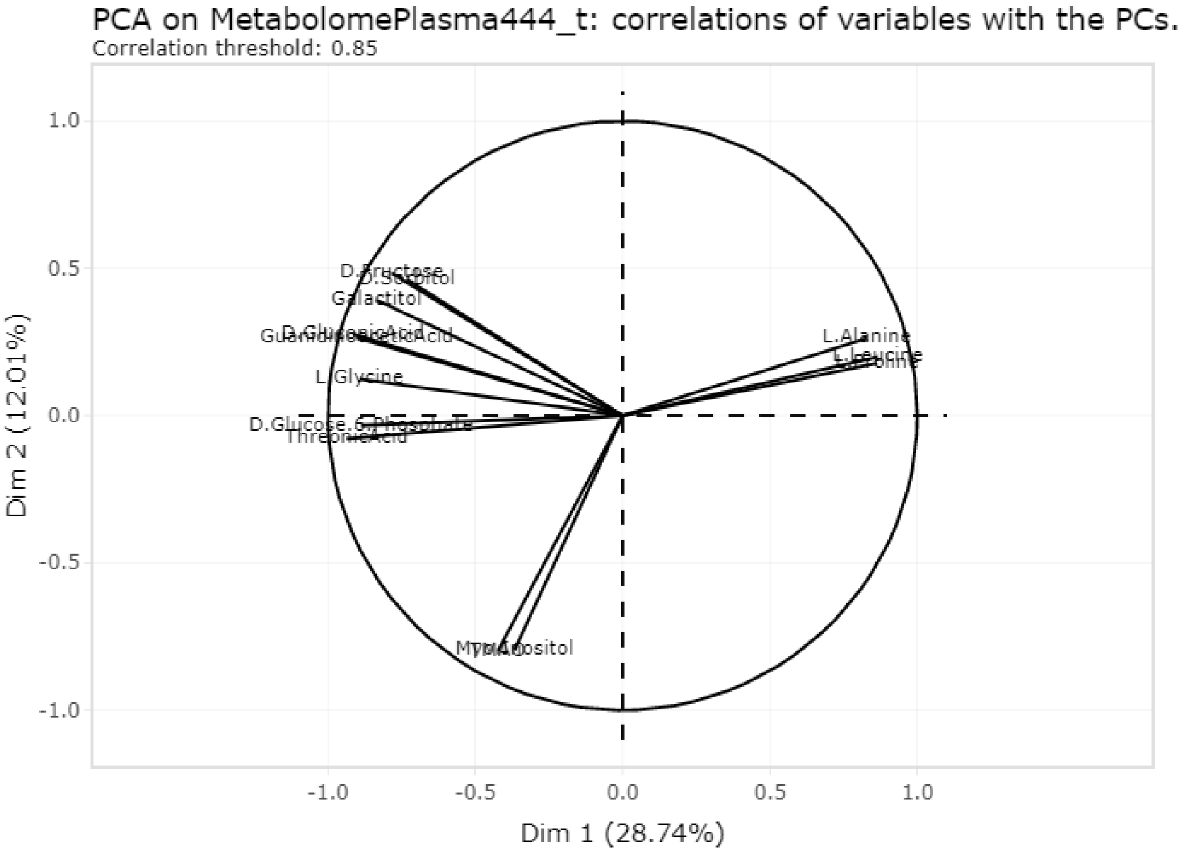


PCA on MetabolomePlasma444_t: projection of individuals on the PC

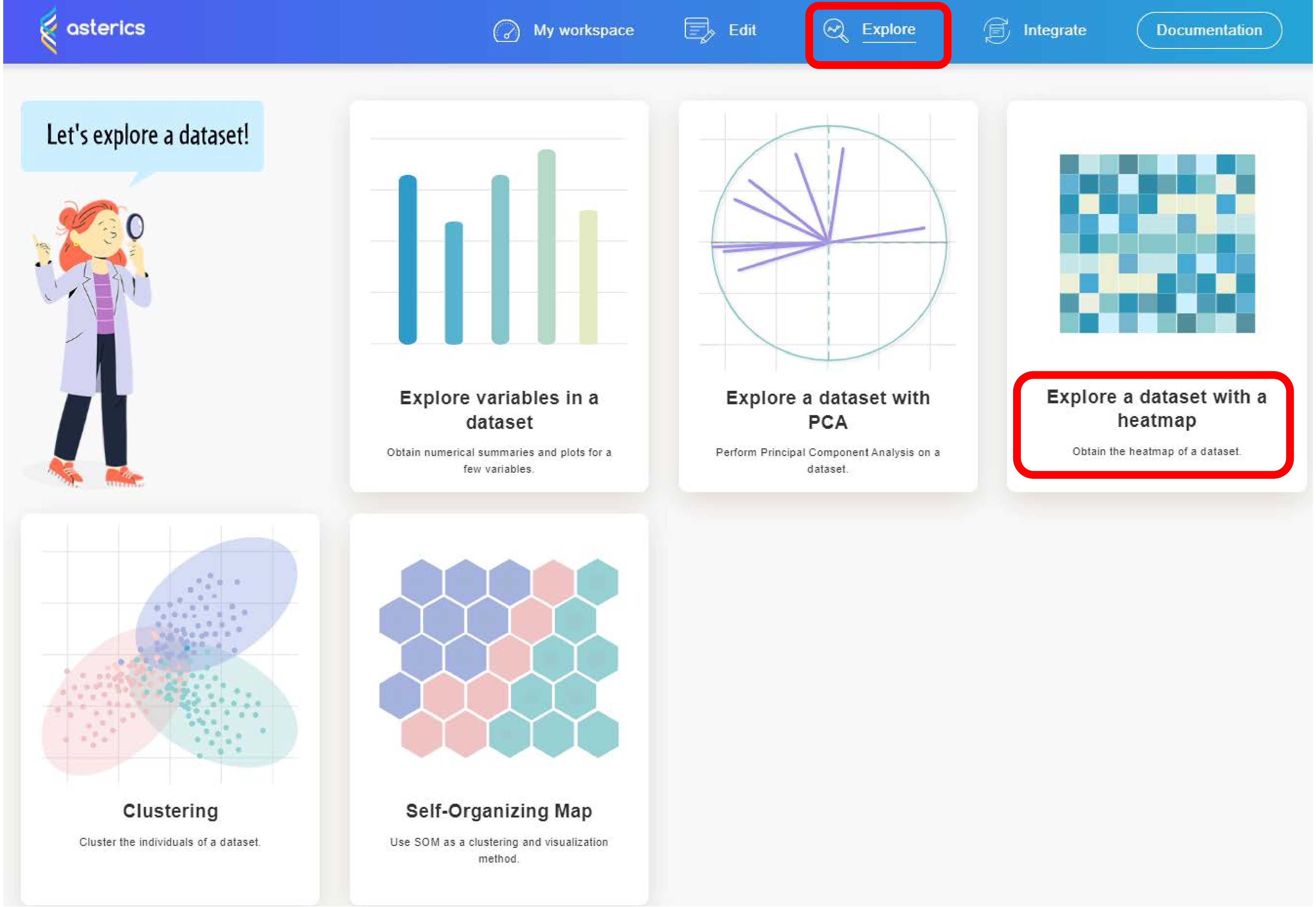


PCA on MetabolomePlasma444_t: projection of individuals on the PC


PCA on MetabolomePlasma444_t: projection of individuals on the PC


PCA on MetabolomePlasma444_t: correlations of variables with the PCs.

Correlation threshold: 0.85



Explore variables in a dataset

Obtain numerical summaries and plots for a few variables.

Explore a dataset with PCA

Perform Principal Component Analysis on a dataset.

Explore a dataset with a heatmap

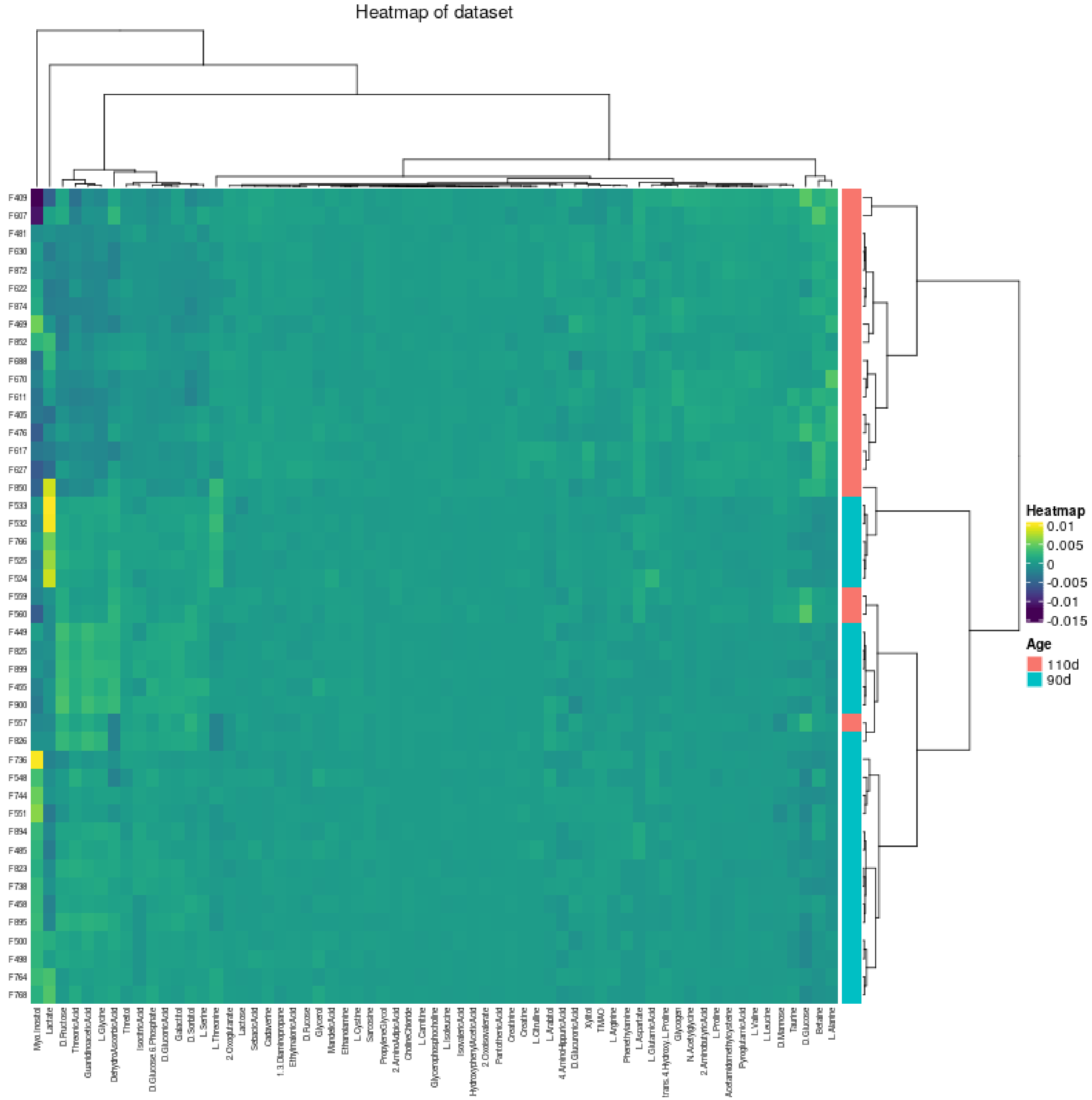
Obtain the heatmap of a dataset.

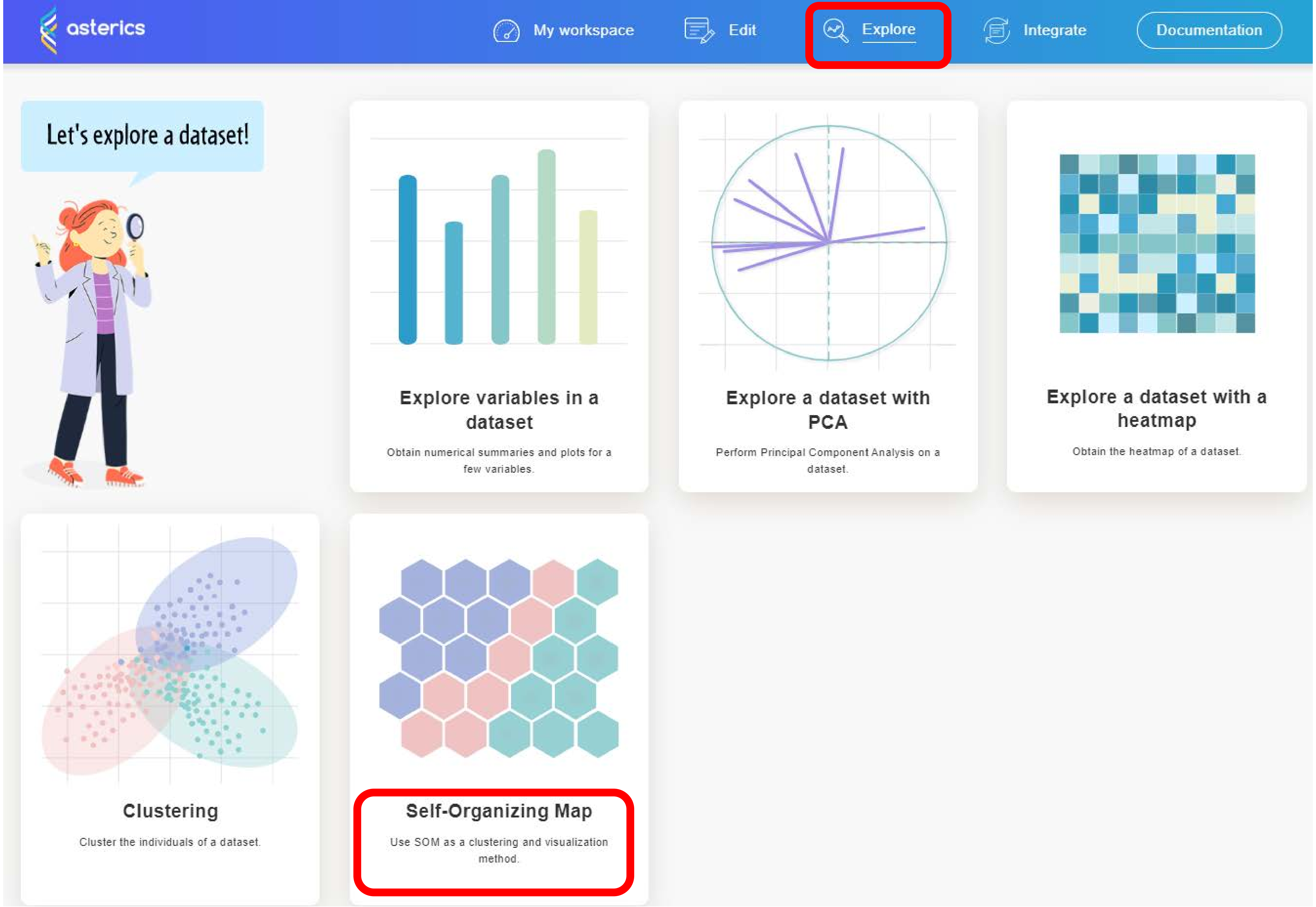
Clustering

Cluster the individuals of a dataset.

Self-Organizing Map

Use SOM as a clustering and visualization method.


Heatmap of dataset
F409
F607
F481
F630
F872
F622
F874
F469
F852
F688
F670
F611
F405
F476
F617
F627
F850
F533
F532
F766
F525
F524
F559
F560
F449
F825
F899
F455
F900
F557
F826
F736
F548
F744
F551
F894
F485
F823
F738
F458
F895
F500
F498
F764
F768
Myo.Inositol
Lactate
D.Fructose
ThreonicAcid
GuanidinoaceticAcid
L.Glycine
DehydroAscorbicAcid
Threitol
IsocitricAcid
D.Glucose.6.Phosphate
D.GluconicAcid
Galactitol
D.Sorbitol
L.Serine
L.Threonine
2.Oxoglutarate
Lactose
SebacicAcid
Cadaverine
1.3.Diaminopropane
EthylmalonicAcid
D.Fucose
Glycerol
MandelicAcid
Ethanolamine
L.Cystine
Sarcosine
PropyleneGlycol
2.AminoAdipicAcid
CholineChloride
L.Carnitine
Glycerophosphocholine
L.Isoleucine
IsovalericAcid
HydroxyphenylAceticAcid
2.Oxoisovalerate
PantothenicAcid
Creatinine
Creatine
L.Citrulline
L.Arabitol
4.AminoHippuricAcid
D.GlucuronicAcid
Xylitol
TMAO
L.Arginine
Phenethylamine
L.Aspartate
L.GlutamicAcid
trans.4.Hydroxy.L.Proline
Glycogen
N.Acetylglycine
2.AminobutyricAcid
L.Proline
Acetamidomethylcysteine
PyroglutamicAcid
L.Valine
L.Leucine
D.Mannose
Taurine
D.Glucose
Betaine
L.Alanine
Heatmap
0.01
0.005
0
-0.005
-0.01
-0.015
Age
110d
90d



Explore variables in a dataset

Obtain numerical summaries and plots for a few variables.

Explore a dataset with PCA

Perform Principal Component Analysis on a dataset.

Explore a dataset with a heatmap

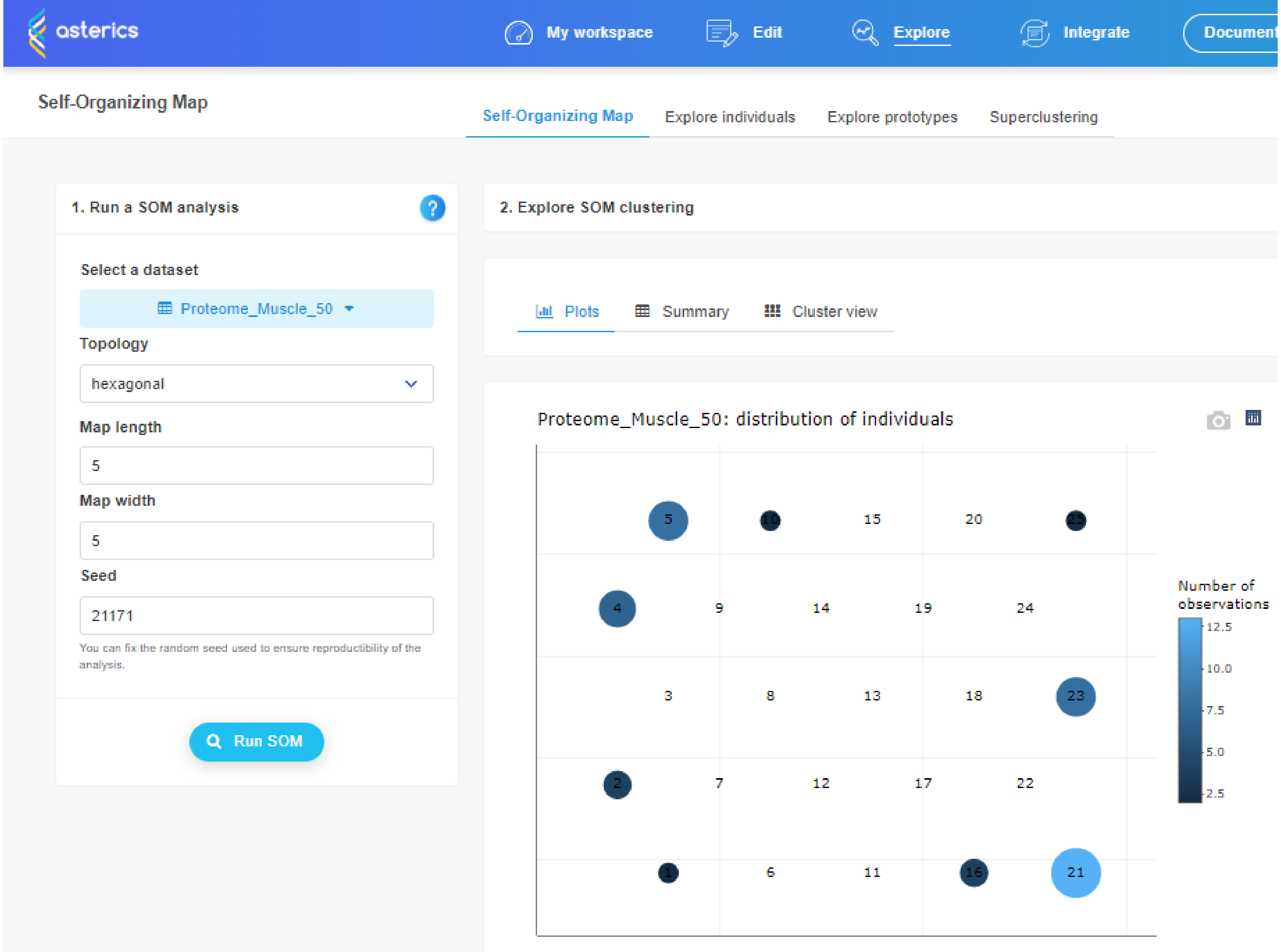
Obtain the heatmap of a dataset.

Clustering

Cluster the individuals of a dataset.

Self-Organizing Map

Use SOM as a clustering and visualization method.



Select a dataset
Proteome_Muscle_50

Topology
hexagonal

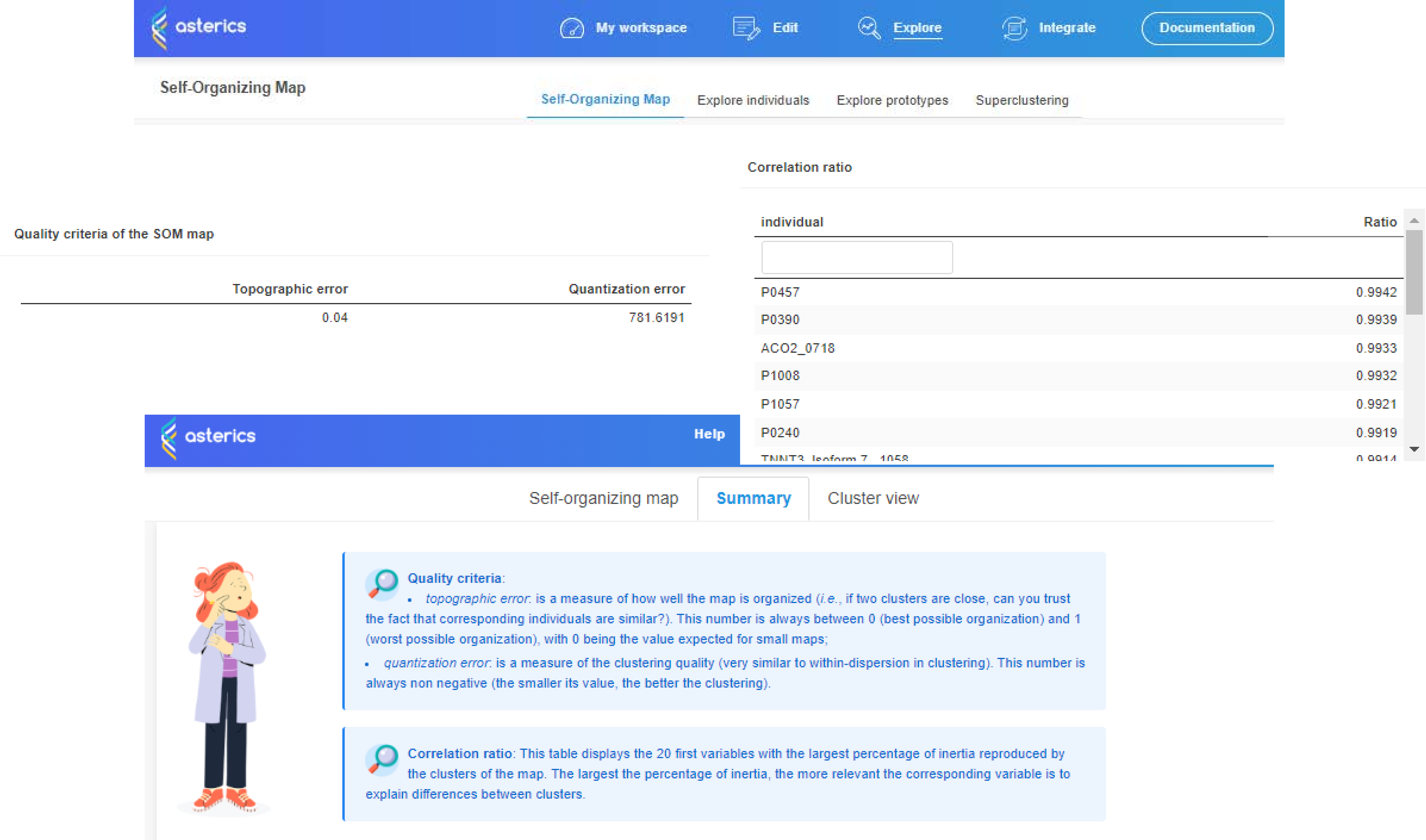
Map length
5

Map width
5

Seed
21171

You can fix the random seed used to ensure reproductibility of the analysis.

Proteome_Muscle_50: distribution of individuals



Quality criteria of the SOM map

Topographic error	Quantization error
0.04	781.6191

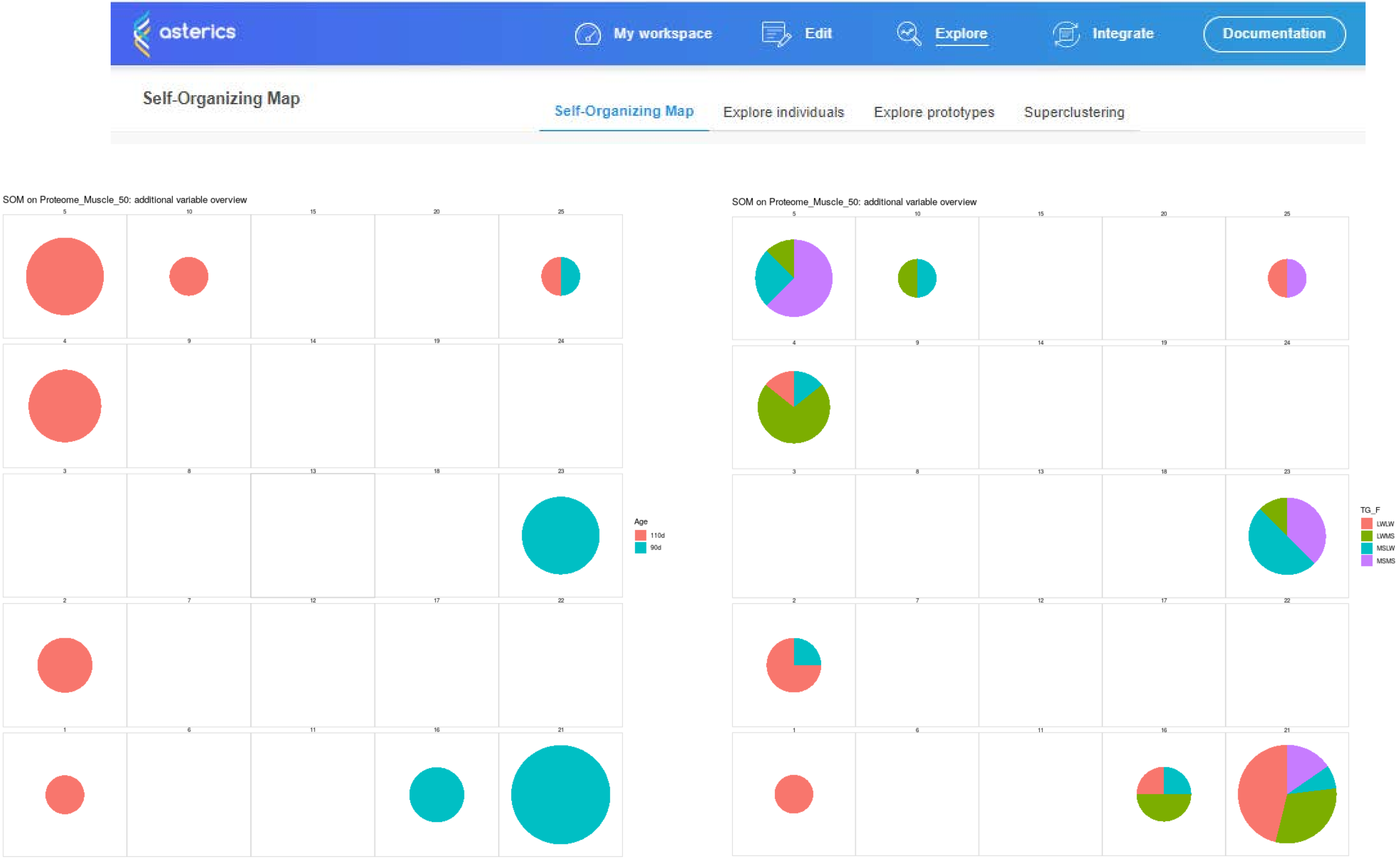
Correlation ratio

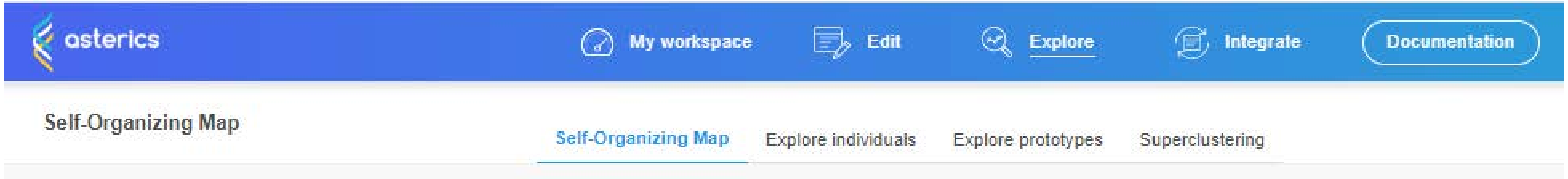
individual	Ratio
P0457	0.9942
P0390	0.9939
ACO2_0718	0.9933
P1008	0.9932
P1057	0.9921
P0240	0.9919
TNNT3_Isoform 7__1058	0.9914

Quality criteria:

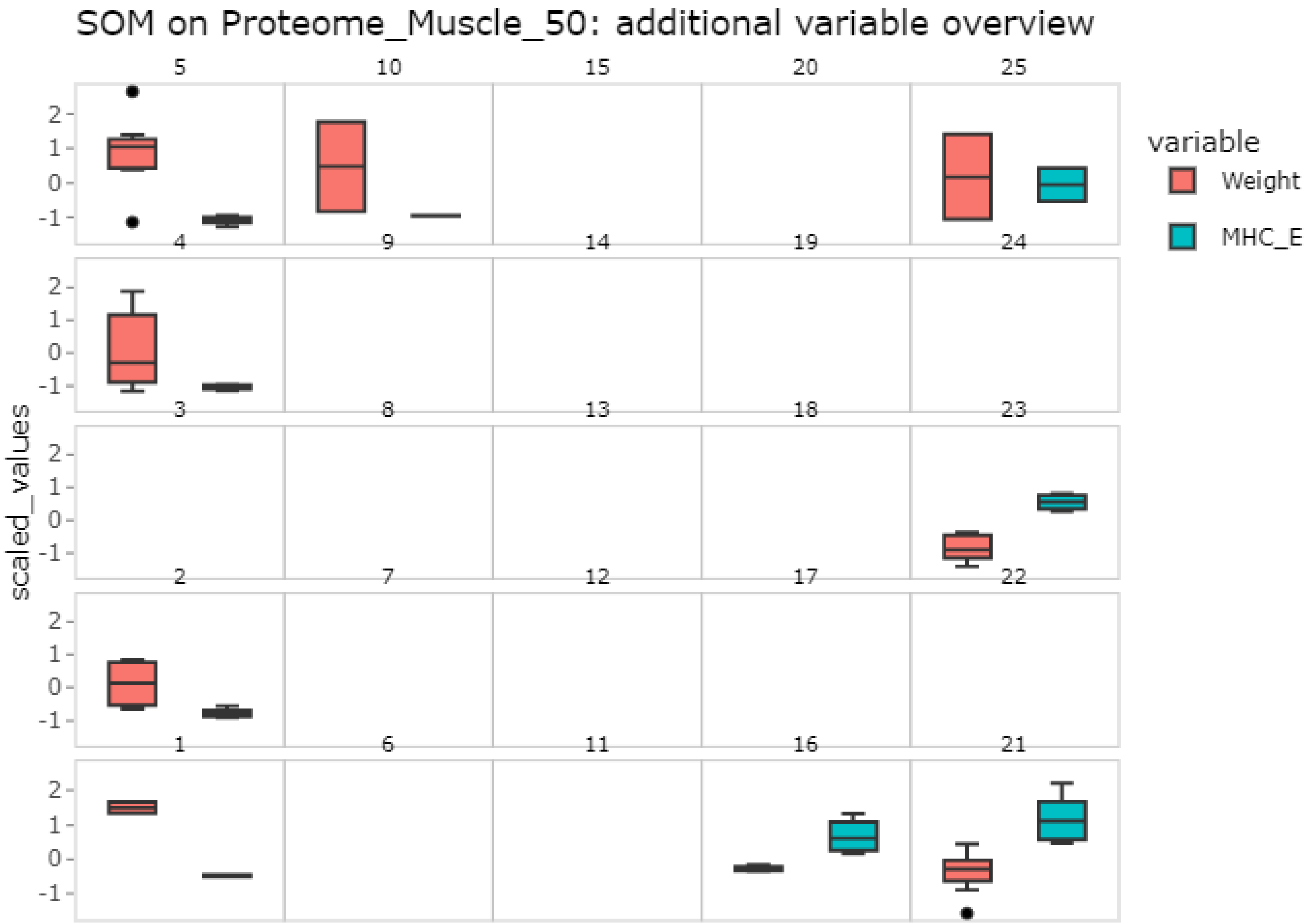
- *topographic error*: is a measure of how well the map is organized (*i.e.*, if two clusters are close, can you trust the fact that corresponding individuals are similar?). This number is always between 0 (best possible organization) and 1 (worst possible organization), with 0 being the value expected for small maps;
- *quantization error*: is a measure of the clustering quality (very similar to within-dispersion in clustering). This number is always non negative (the smaller its value, the better the clustering).

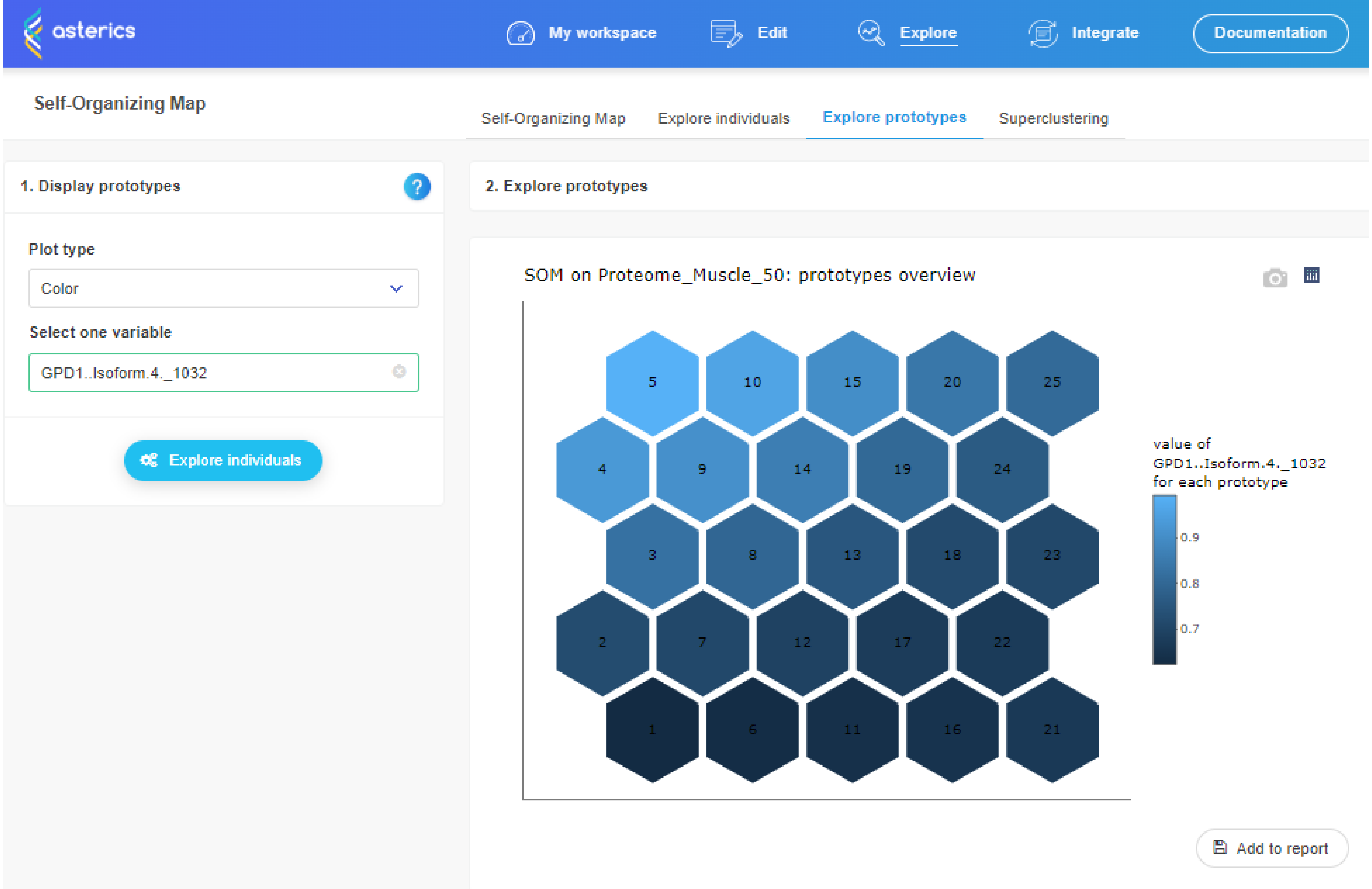
Correlation ratio: This table displays the 20 first variables with the largest percentage of inertia reproduced by the clusters of the map. The largest the percentage of inertia, the more relevant the corresponding variable is to explain differences between clusters.


asterics
My workspace
Edit
Explore
Integrate
Documentation
Self-Organizing Map
Self-Organizing Map
Explore individuals
Explore prototypes
Superclustering
SOM on Proteome_Muscle_50: additional variable overview
5
10
15
20
25
4
9
14
19
24
3
8
13
18
23
2
7
12
17
22
1
6
11
16
21
Age
110d
90d
SOM on Proteome_Muscle_50: additional variable overview
5
10
15
20
25
4
9
14
19
24
3
8
13
18
23
2
7
12
17
22
1
6
11
16
21
TG_F
LWLW
LWMS
MSLW
MSMS


asterics
My workspace
Edit
Explore
Integrate
Documentation
Self-Organizing Map
Self-Organizing Map
Explore individuals
Explore prototypes
Superclustering


SOM on Proteome_Muscle_50: additional variable overview
Age
110d
90d


SOM on Proteome_Muscle_50: additional variable overview
scaled_values
5
10
15
20
25
4
9
14
19
24
3
8
13
18
23
2
7
12
17
22
1
6
11
16
21
2
1
0
-1
variable
Weight
MHC_E



Self-Organizing Map

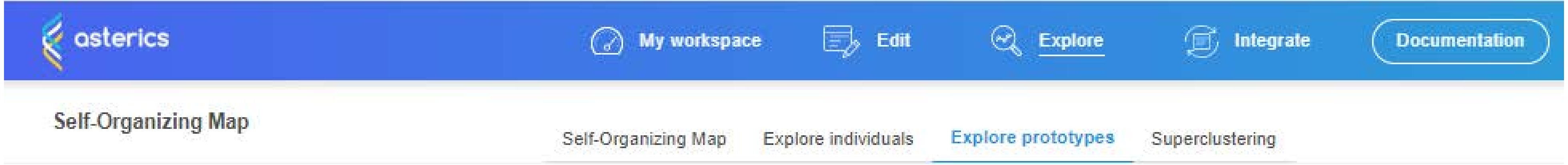
1. Display prototypes

Plot type: Color

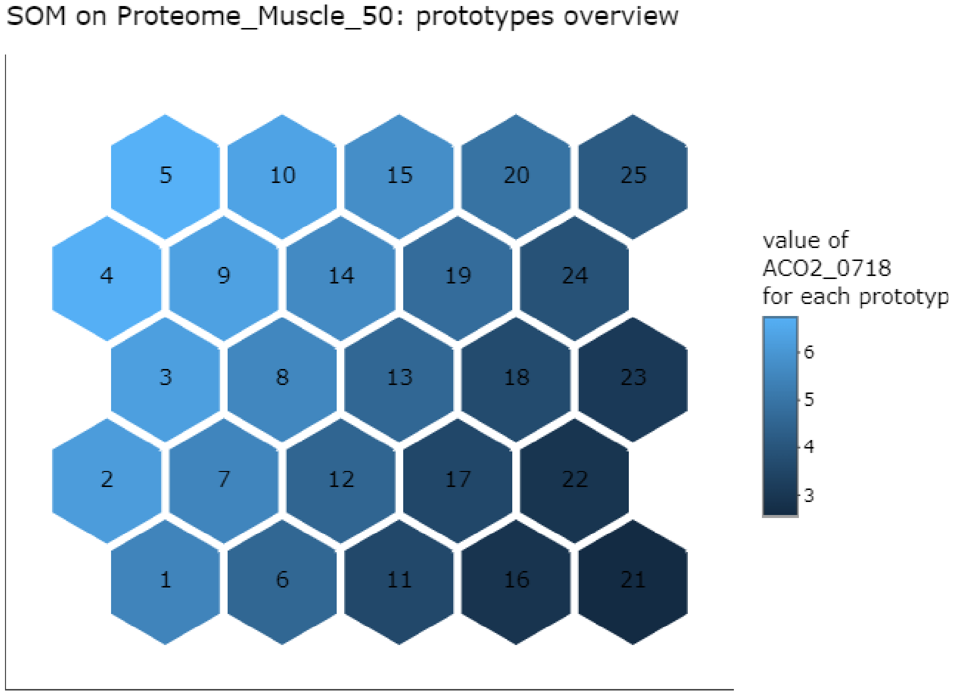
Select one variable: GPD1..Isoform.4._1032

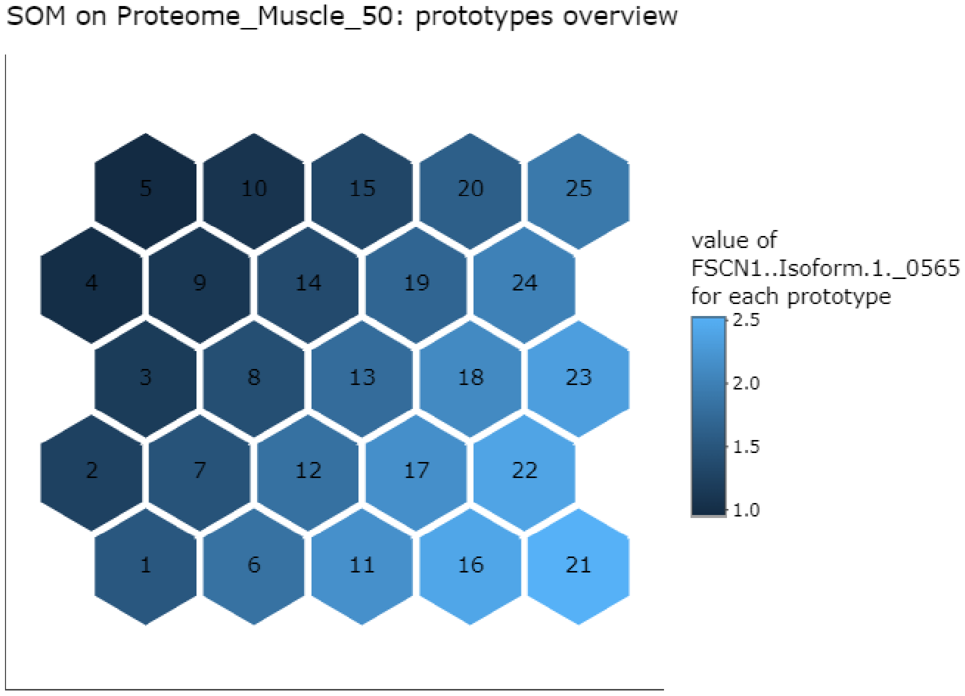
2. Explore prototypes

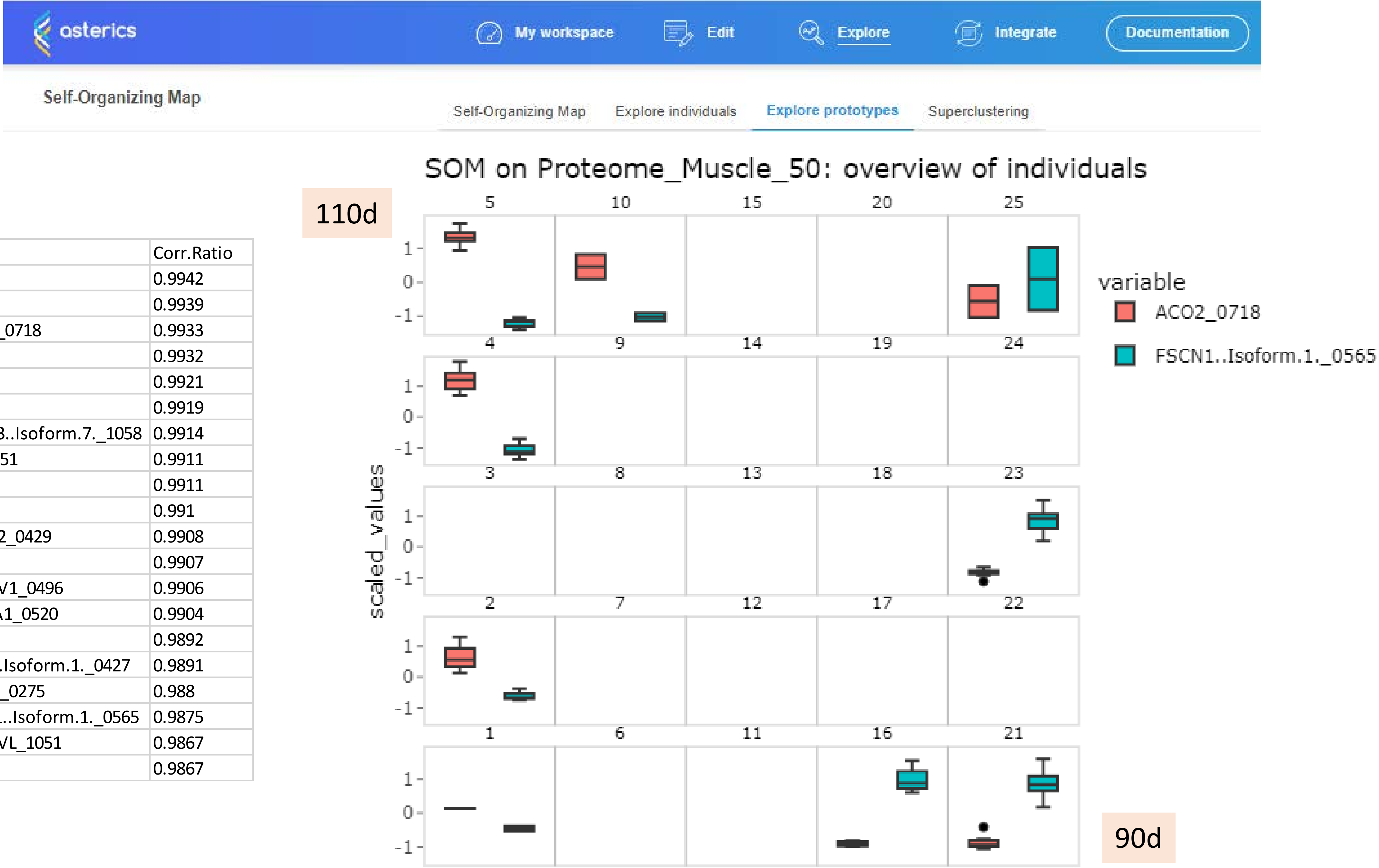
SOM on Proteome_Muscle_50: prototypes overview



	Corr.Ratio
P0457	0.9942
P0390	0.9939
ACO2_0718	0.9933
P1008	0.9932
P1057	0.9921
P0240	0.9919
TNNT3..Isoform.7._1058	0.9914
FH_0451	0.9911
P0913	0.9911
P0773	0.991
CKMT2_0429	0.9908
P0719	0.9907
NDUFV1_0496	0.9906
ATP5A1_0520	0.9904
P0269	0.9892
PGK1..Isoform.1._0427	0.9891
MDH2_0275	0.988
FSCN1..Isoform.1._0565	0.9875
ACADVL_1051	0.9867
P0276	0.9867


SOM on Proteome_Muscle_50: prototypes overview

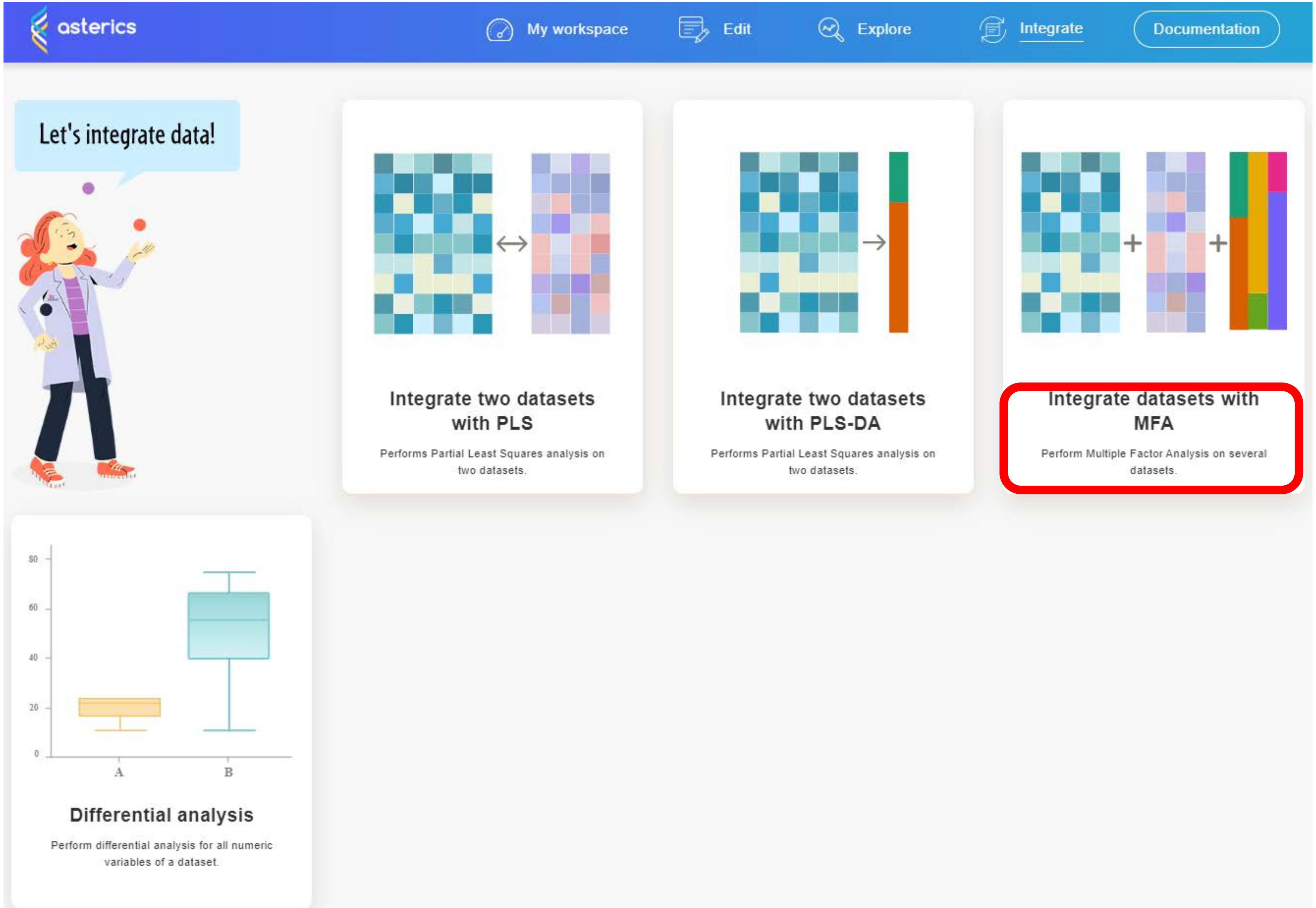

SOM on Proteome_Muscle_50: prototypes overview



	Corr.Ratio
P0457	0.9942
P0390	0.9939
ACO2_0718	0.9933
P1008	0.9932
P1057	0.9921
P0240	0.9919
TNNT3..Isoform.7._1058	0.9914
FH_0451	0.9911
P0913	0.9911
P0773	0.991
CKMT2_0429	0.9908
P0719	0.9907
NDUFV1_0496	0.9906
ATP5A1_0520	0.9904
P0269	0.9892
PGK1..Isoform.1._0427	0.9891
MDH2_0275	0.988
FSCN1..Isoform.1._0565	0.9875
ACADVL_1051	0.9867
P0276	0.9867

110d

90d



Integrate two datasets with PLS

Performs Partial Least Squares analysis on two datasets.

Integrate two datasets with PLS-DA

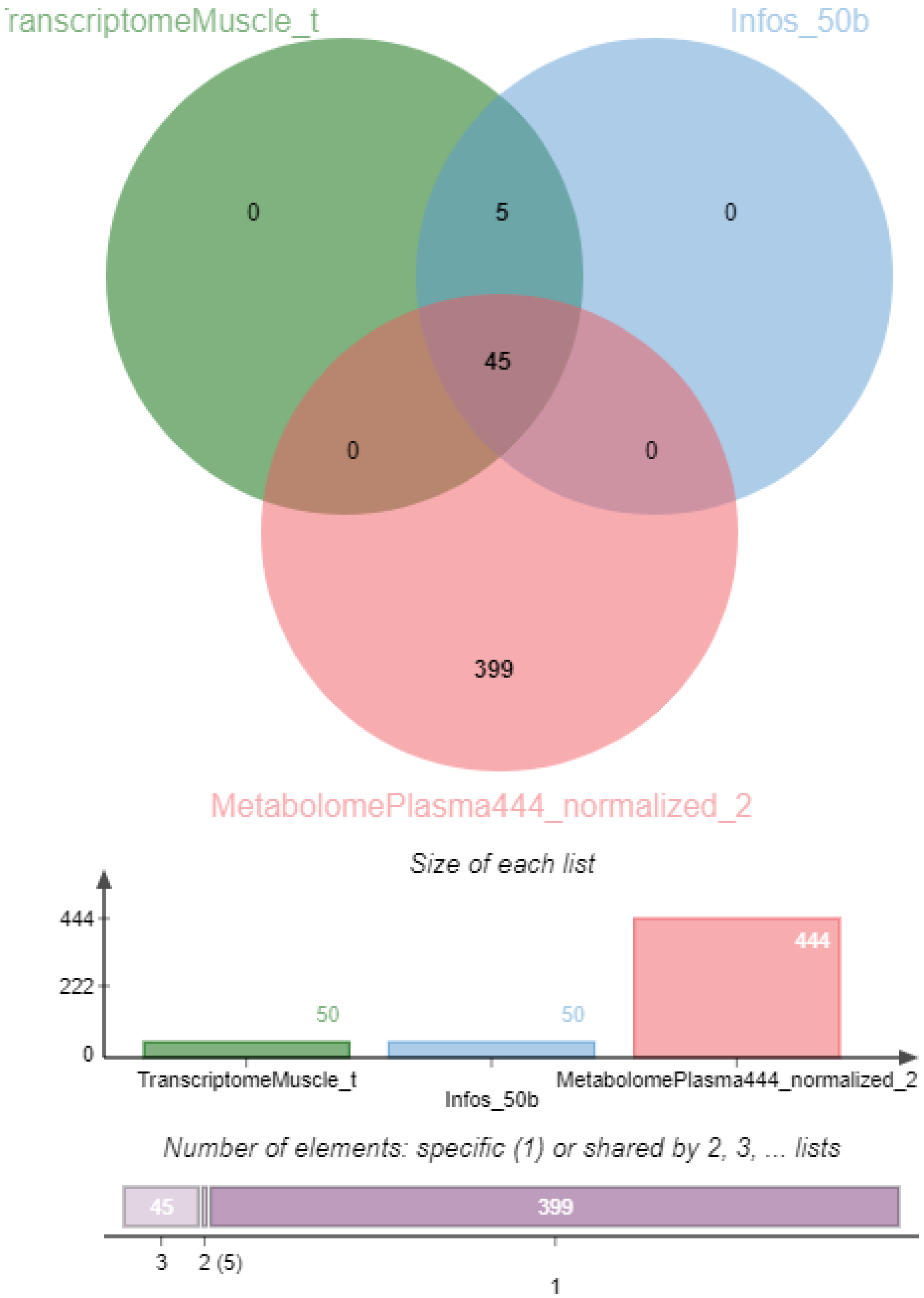
Performs Partial Least Squares analysis on two datasets.

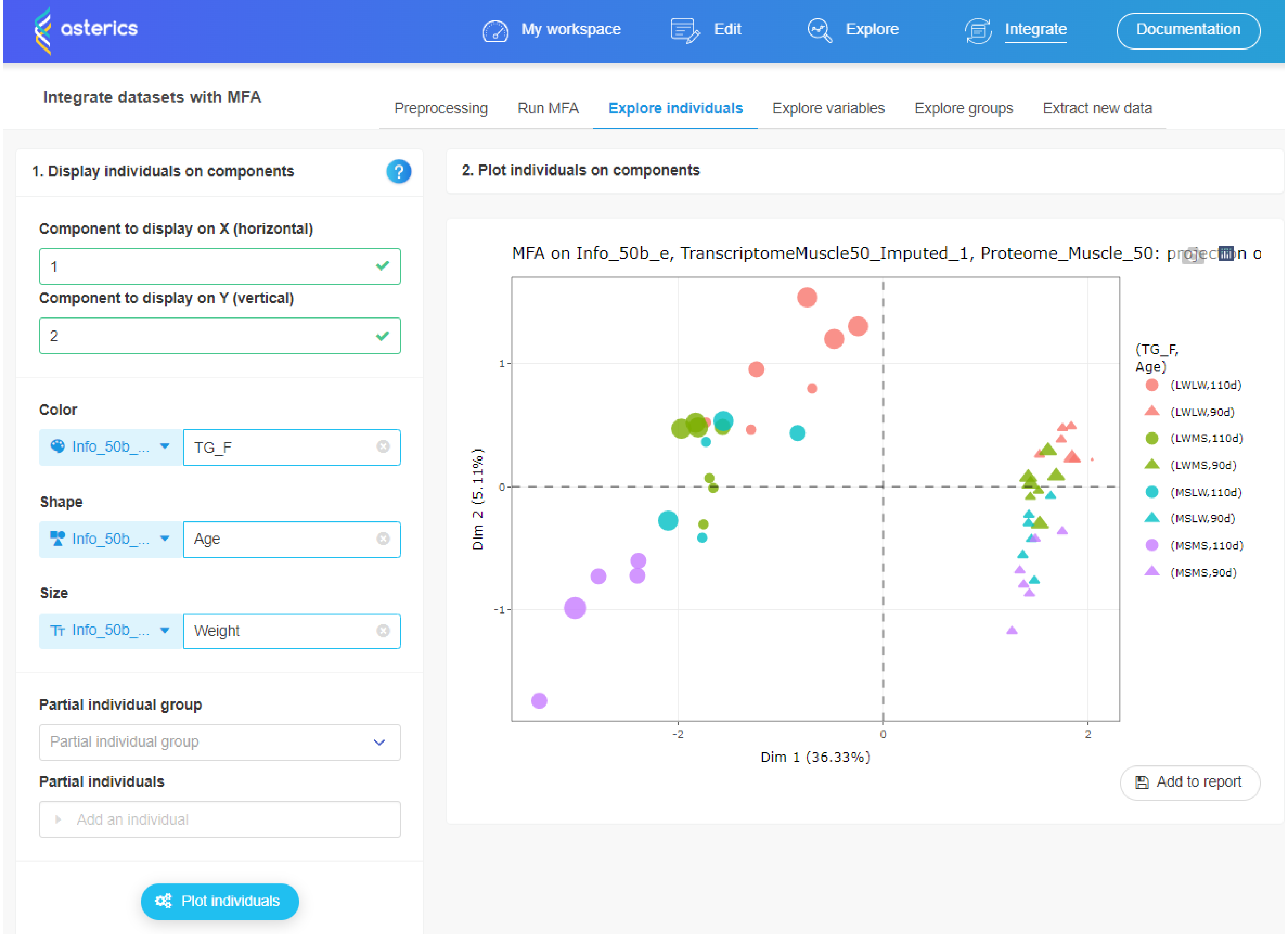
Integrate datasets with MFA

Perform Multiple Factor Analysis on several datasets.

Differential analysis

Perform differential analysis for all numeric variables of a dataset.


ranscriptomeMuscle_t
Infos_50b
0
5
0
45
0
0
399
MetabolomePlasma444_normalized_2
Size of each list
444
222
0
50
50
444
TranscriptomeMuscle_t
Infos_50b
MetabolomePlasma444_normalized_2
Number of elements: specific (1) or shared by 2, 3, ... lists
45
399
3
2 (5)
1



Integrate datasets with MFA

1. Display individuals on components

Component to display on X (horizontal): 1

Component to display on Y (vertical): 2

Color: Info_50b_... TG_F

Shape: Info_50b_... Age

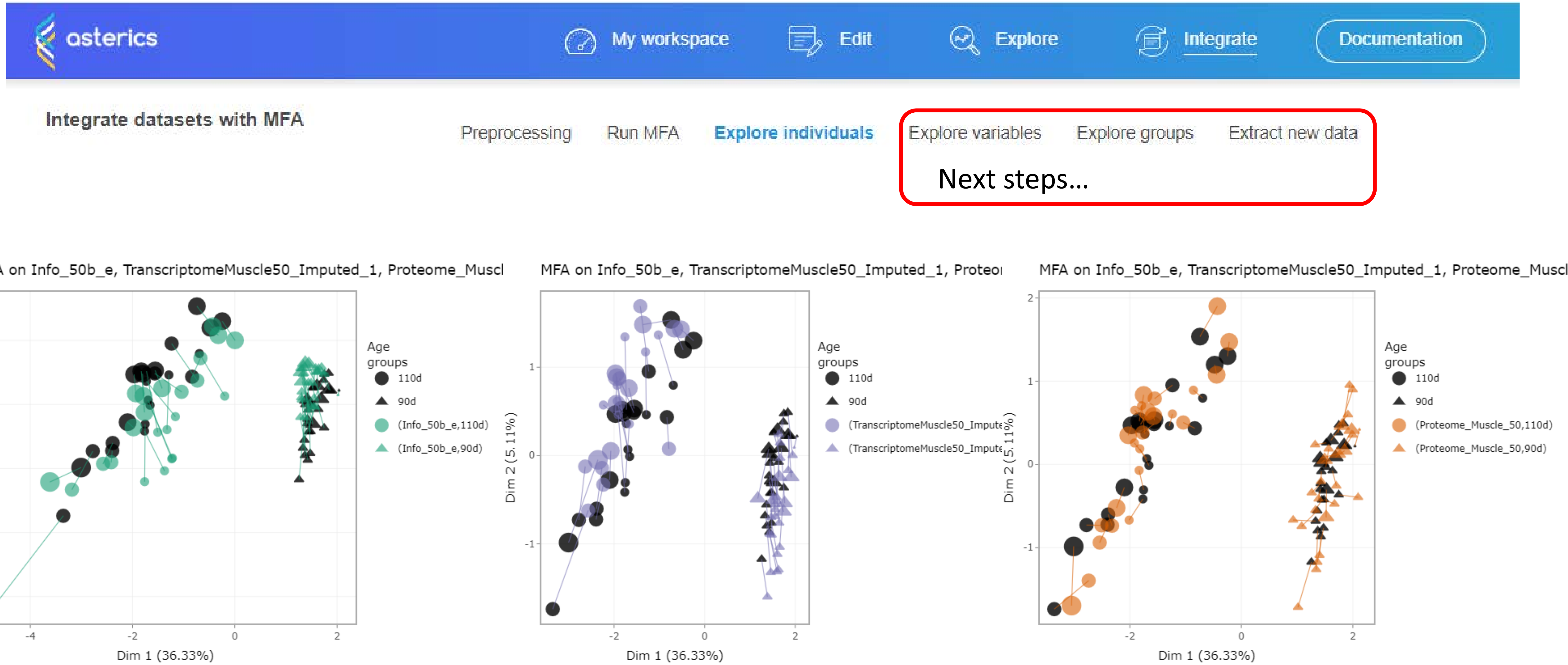
Size: Info_50b_... Weight

Partial individual group

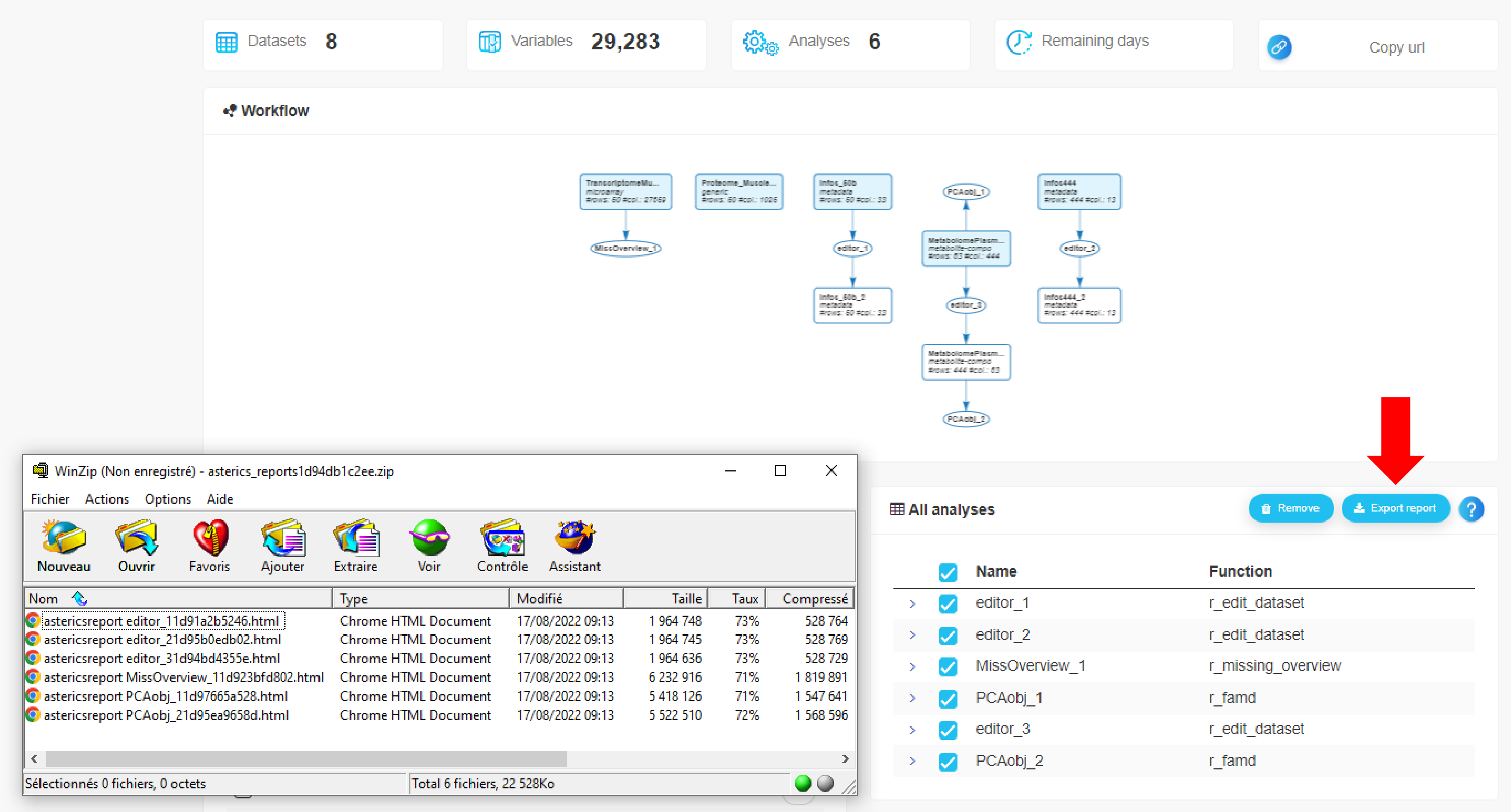
Partial individuals

2. Plot individuals on components

MFA on Info_50b_e, TranscriptomeMuscle50_Imputed_1, Proteome_Muscle_50: projection o


asterics
My workspace
Edit
Explore
Integrate
Documentation
Integrate datasets with MFA
Preprocessing
Run MFA
Explore individuals
Explore variables
Explore groups
Extract new data
MFA on Info_50b_e, TranscriptomeMuscle50_Imputed_1, Proteome_Muscl
Dim 1 (36.33%)
Dim 2 (5.11%)
Age groups
110d
90d
(Info_50b_e,110d)
(Info_50b_e,90d)
MFA on Info_50b_e, TranscriptomeMuscle50_Imputed_1, Proteo
Dim 1 (36.33%)
Dim 2 (5.11%)
Age groups
110d
90d
(TranscriptomeMuscle50_Impute
(TranscriptomeMuscle50_Impute
MFA on Info_50b_e, TranscriptomeMuscle50_Imputed_1, Proteome_Muscl
Dim 1 (36.33%)
Dim 2 (5.11%)
Age groups
110d
90d
(Proteome_Muscle_50,110d)
(Proteome_Muscle_50,90d)

Next steps...



Nom	Type	Modifié	Taille	Taux	Compressé
astericsreport editor_11d91a2b5246.html	Chrome HTML Document	17/08/2022 09:13	1 964 748	73%	528 764
astericsreport editor_21d95b0edb02.html	Chrome HTML Document	17/08/2022 09:13	1 964 745	73%	528 769
astericsreport editor_31d94bd4355e.html	Chrome HTML Document	17/08/2022 09:13	1 964 636	73%	528 729
astericsreport MissOverview_11d923bfd802.html	Chrome HTML Document	17/08/2022 09:13	6 232 916	71%	1 819 891
astericsreport PCAobj_11d97665a528.html	Chrome HTML Document	17/08/2022 09:13	5 418 126	71%	1 547 641
astericsreport PCAobj_21d95ea9658d.html	Chrome HTML Document	17/08/2022 09:13	5 522 510	72%	1 568 596

Sélectionnés 0 fichiers, 0 octets

Total 6 fichiers, 22 528Ko

	☑	Name	Function
›	☑	editor_1	r_edit_dataset
›	☑	editor_2	r_edit_dataset
›	☑	MissOverview_1	r_missing_overview
›	☑	PCAobj_1	r_famd
›	☑	editor_3	r_edit_dataset
›	☑	PCAobj_2	r_famd

Merci à toute l'équipe d'Asterics
Merci pour votre attention

To be continued…

asterics
A tool for the exploration and integration of omics data