

Step-by-step Guide to Preprocess Expression Data using FlexStat Pipeline - Sample data



This feature facilitates preprocessing expression data with experimentally generated data. It involves missing value imputation and data normalization where users can specify the algorithm, and method to be used.

This tutorial is based on sample data into the application.

1

Navigate to <https://jglab.shinyapps.io/flexstatv1-pipeline-only/>

2

Go to "Data Preparation" tab.

3 Click "Use Sample Data"

FlexStat 1.0 Data Preparation Differential Expression Automated Combinatory Differential Expression Consensus C

Data Preprocessing

Select CSV File to Import ⓘ

Browse... No file selected

☒ Show head ☒ Use Sample Data ⓘ

☐ Transpose data ☐ Log2 Transform ☐ Log10 Transform

Select columns to remove

Class Variable

Not Selected

Data Preprocessed Data Normalized

Sample Data

Condition	Batch	O76070	P01344
A	Set1	28.41	27.36
A	Set1	28.46	27.40
A	Set1	28.41	27.47
B	Set2	24.28	24.63
B	Set2	24.28	24.73
B	Set2	24.20	24.66

4 Select 'Class' variable to generate quality control plots.

Browse... No file selected

☒ Show head ☒ Use Sample Data ⓘ

☐ Transpose data ☐ Log2 Transform ☐ Log10 Transform

Select columns to remove

Class Variable

Not Selected

Not Selected

Condition

Batch

Select Data Normalization Method

☒ Median Normalization ☐ Quantile Normalization ☐ Internal Reference Normalization

Missing value threshold

Sample Data

Condition	Batch	O76070	P01344
A	Set1	28.41	27.36
A	Set1	28.46	27.40
A	Set1	28.41	27.47
B	Set2	24.28	24.63
B	Set2	24.28	24.73
B	Set2	24.20	24.66

5 Select "Experimental batch variable" to be used in normalization.

☐ Transpose data ☐ Log2 Transform ☐ Log10 Transform

Select columns to remove

Class Variable

Condition

Experimental Batch Variable

Not Selected

Select Data Normalization Method

☒ Median Normalization

☐ Quantile Normalization

☐ Internal Reference Normalization

Missing value threshold

0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

A	Set1	28.41	27.36
A	Set1	28.46	27.40
A	Set1	28.41	27.47
B	Set2	24.28	24.63
B	Set2	24.28	24.73
B	Set2	24.20	24.66

6 Select missing value threshold. Default is 0.5 which means proteins with 50% or less missing values will be considered for imputation.

Missing value threshold

0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

7 Select Data Normalization method

1. Median normalization
2. Quantile normalization
3. Internal Reference Normalization: Recommended to use for label-based data
4. if selected, select the corresponding internal reference to be used
5. Variance Stabilization Normalization: Recommended to use for label-free data

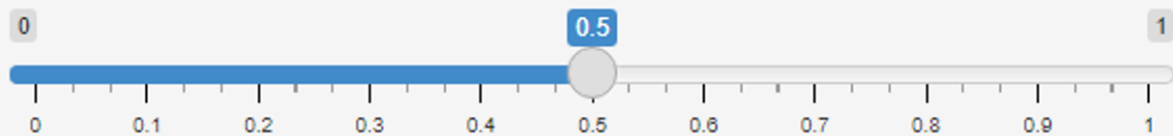
Experimental Batch Variable

Batch

Select Data Normalization Method

- ☒ Median Normalization
- ☐ Quantile Normalization
- ☐ Internal Reference Normalization
- ☐ Variance Stabilization Normalization

Missing value threshold



8

Internal Reference Normalization: if selected, then select the corresponding internal reference to be used.

Select the corresponding internal reference to be used. Here we select "A1 (log)" and "B1 (log)".

Select Data Normalization Method

- ☐ Median Normalization
- ☐ Quantile Normalization
- ☒ Internal Reference Normalization

Select internal references

A1 (log) B1 (log)

A2 (log)

A3 (log)

B2 (log)

B3 (log)

Select Data Imputation Method

- ☒ Random draw from a normal distribution
- ☐ K-nearest neighbour
- ☐ MissForest

Preprocess Data

9

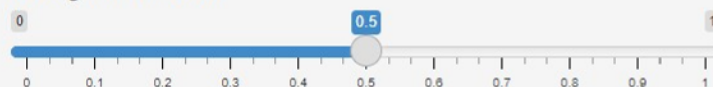
Click "Preprocess Data"

- ☐ Median Normalization
- ☐ Quantile Normalization
- ☒ Internal Reference Normalization

Select internal references

A1 (log) B1 (log)

Missing value threshold



Select Data Imputation Method

- ☒ Random draw from a normal distribution
- ☐ K-nearest neighbour
- ☐ MissForest

Preprocess Data

- 10 Click "Preprocessed Data" to check the preprocessing results.

Differential Expression Automated Combinatory Differential Expression Consensus Clustering Help

Data **Preprocessed Data** Normalization Quality Control Plots Imputation Quality Control Plots

☒ Use Sample Data ⓘ

orm ☐ Log10 Transform

Sample Data

Condition	Batch	O76070	P01344	P01579	P00709	P41159	P00918
A	Set1	28.41	27.36	27.40	27.14	28.23	28.04
A	Set1	28.46	27.40	27.37	27.05	28.14	28.07
A	Set1	28.41	27.47	27.37	27.08	28.23	28.03
B	Set2	24.28	24.63	23.63	22.84	24.47	24.17
B	Set2	24.28	24.73	23.44	23.03	24.72	24.52
B	Set2	24.20	24.66	23.68	22.76	24.66	24.47

- 11 Click "Download Normalized Data" to download the normalized data.

Differential Expression Automated Combinatory Differential Expression Consensus Clustering Help

Data Preprocessed Data **Normalization Quality Control Plots** Imputation Quality Control Plots

☒ Use Sample Data ⓘ

☐ Log10 Transform

Normalized Data

Show entries **Download Normalized Data**

	O76070	P01344	P01579	P00709	P41159	P00918	P01112
A1 (log)	26.3878	26.0434	25.5596	25.0298	26.3950	26.1504	25.249
A2 (log)	26.4464	26.0915	25.5373	24.9584	26.3320	26.1846	25.450
A3 (log)	26.3943	26.1573	25.5392	24.9800	26.4044	26.1496	25.444

12 Click "Download Imputed Data" to download the imputed data.

Showing 1 to 6 of 6 entries

Imputed Data

Show 10 entries [Download Imputed Data](#)

	O76070	P01344	P01579	P00709	P41159	P00918	P01112	C
B2 (log)	26.2183	25.9773	25.1872	25.0684	26.4885	26.3558	25.4090	
B3 (log)	26.1499	25.9165	25.4601	24.7902	26.4399	26.3178	25.3551	
A1 (log)	26.3878	26.0434	25.5596	25.0298	26.3950	26.1504	25.2496	
A2 (log)	26.4464	26.0915	25.5373	24.9584	26.3320	26.1846	25.4503	
A3 (log)	26.3943	26.1573	25.5392	24.9800	26.4044	26.1496	25.4440	

13 Click "Normalization Quality Control Plots" to view the outputs of normalization steps.

Use Sample Data ⓘ

Log10 Transform

Data Preprocessed Data **Normalization Quality Control Plots** Imputation Quality Control Plots

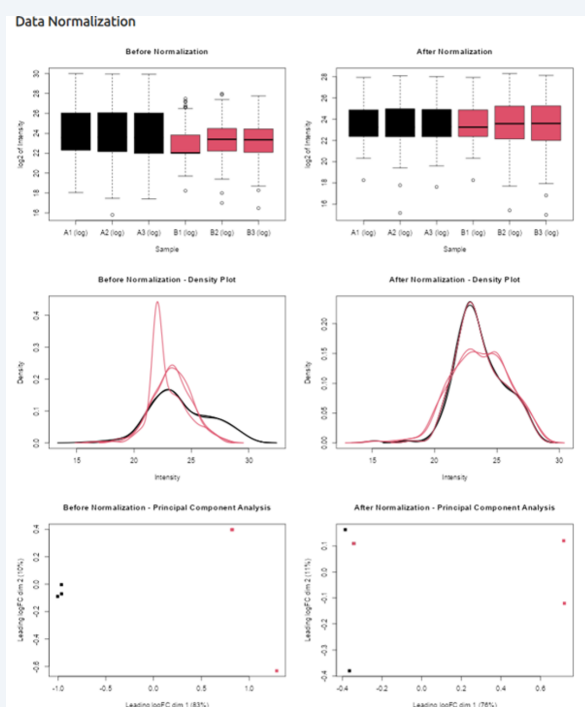
Normalized Data

Show 10 entries [Download Normalized Data](#)

	O76070	P01344	P01579	P00709	P41159	P00918	P01112	Q158
A1 (log)	26.3878	26.0434	25.5596	25.0298	26.3950	26.1504	25.2496	23.1
A2 (log)	26.4464	26.0915	25.5373	24.9584	26.3320	26.1846	25.4503	23.1
A3 (log)	26.3943	26.1573	25.5392	24.9800	26.4044	26.1496	25.4440	23.1

14

Click "Normalization Quality Control Plots" to view the outputs of normalization steps.



15

Click "Imputation Quality Control Plots"



16 Click "Imputation Quality Control Plots"

