

MetNormalizer Tutorial

1. Installation of MetNormalizer

(1) First install R program (<http://www.r-project.org/>), since MetNormalizer is developed with R.

(2) Install R package e1071 via R command:

```
install.packages ("e1071")
```

(3) Install the MetNormalizer Package in R.

Place the R packages MetNormalizer_1.0.tar.gz and MetNormalizer_1.0.zip in a folder, and set this folder as directory:

```
setwd ('/path/to/your/packages')
```

a) For Windows, install the MetNormalizer Package via R command:

```
install.packages ("MetNormalizer_1.0.zip", repos=NULL, type="source")
```

b) For Linux and Mac OS, install the MetNormalizer Package via R command:

```
install.packages ("MetNormalizer_1.0.tar.gz", repos=NULL, type="source")
```

2. Data Analysis

Step 1: Prepare data files for data analysis

Create a folder named as "normalization"; organize your dataset as three csv files, name as "qc.csv", "sample.csv", and "tags.csv"; and put three csv files into the "normalization" folder:



1.1 "qc.csv" – provide the raw data, name, and injection order of QC samples. The second row of the table must be the injection order of QC samples. The example for the "qc.csv" is given below:

	QC1	QC2	QC3	QC4	QC5	QC name
	2	3	4	5	6	Injection order
1	686045.4	711687.1	730865.6	609289.8	633877.4	intensity
2	155214.5	157294.4	147452	134176	125833.2	
3	42182.63	38494.04	36199.81	35701.75	33145.6	
4	174746.5	166427.5	171377.9	160315	152272.1	
5	104626.8	77539.47	78756.58	71194.05	68859.18	
6	530059.6	384550.9	355183.8	629222.5	458856	

feature name

1.2 “sample.csv” – provide the raw data, name and injection order of subject samples. The second row of the table must be the injection order. The example for the “sample.csv” is given below:

	S20	S30	S39	S49	S59
	2	3	4	5	6
1	706899.2	777526.7	750672.7	684709.1	719189.2
2	159022.8	179071.1	167932.1	175157.5	168632.9
3	39347.48	40378.69	40705.77	41078	40361.4
4	172586.4	169093.5	173186.1	177013.2	169348.6
5	84064.91	102886.4	89647.29	95198.7	86546.9
6	261829.5	516845.6	487036	243869.5	285061.9

Annotations: A blue arrow points to the header row (S20-S59) labeled "sample name". A black arrow points to the second row (2-6) labeled "Injection order". A red bracket on the left side of the data rows is labeled "feature name". A purple bracket on the right side of the data rows is labeled "intensity".

1.3 “tags.csv” – provide the mass spectrometry information of features, for example, m/z (mass-to-charge ratio), rt (retention time), metabolite ID, metabolite name and so on. **It at least contains two columns for "mz" and "rt"**. The example for the “tags.csv” is given below:

feature name	mz	rt
1	76.07598	517.004
2	83.06036	473.104
3	85.02796	435.884
4	86.07103	240.986
5	104.0701	417.792
6	104.9922	729.739

A blue bracket above the table is labeled "feature information".

Step 2. Process the data files

2.1 Load MetNormalizer R package via the following command:

```
library ( MetNormalizer)
```

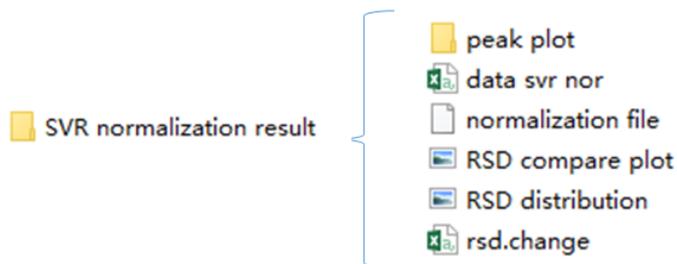
2.2 Set directory of data files:

```
setwd('/path/to/your/data/normalization')
```

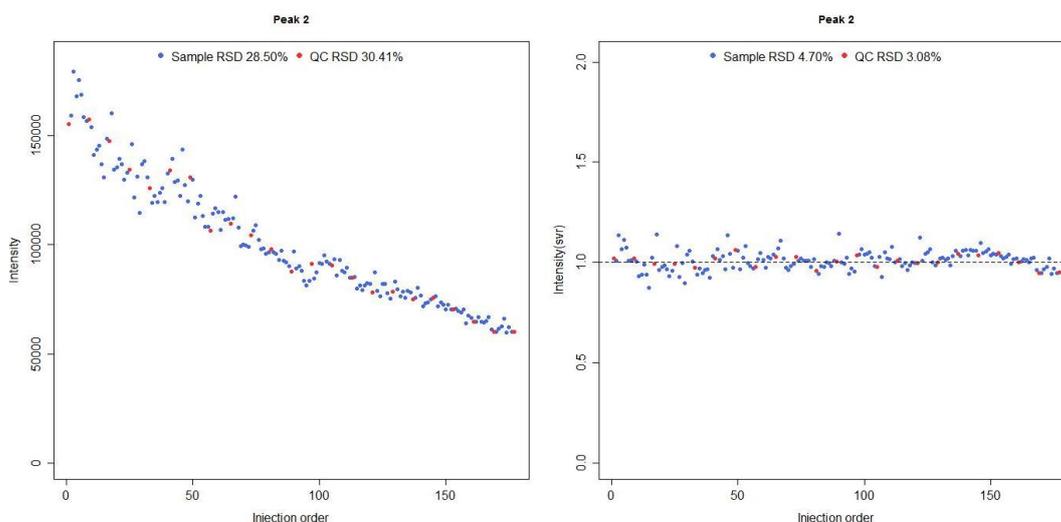
2.3 Process the data:

```
MetNormalizer ()
```

2.4 Results are automatically output after the completion of data processing. All the output files are placed in a new folder named “SVR normalization result”. The explanations for each output file are given below:



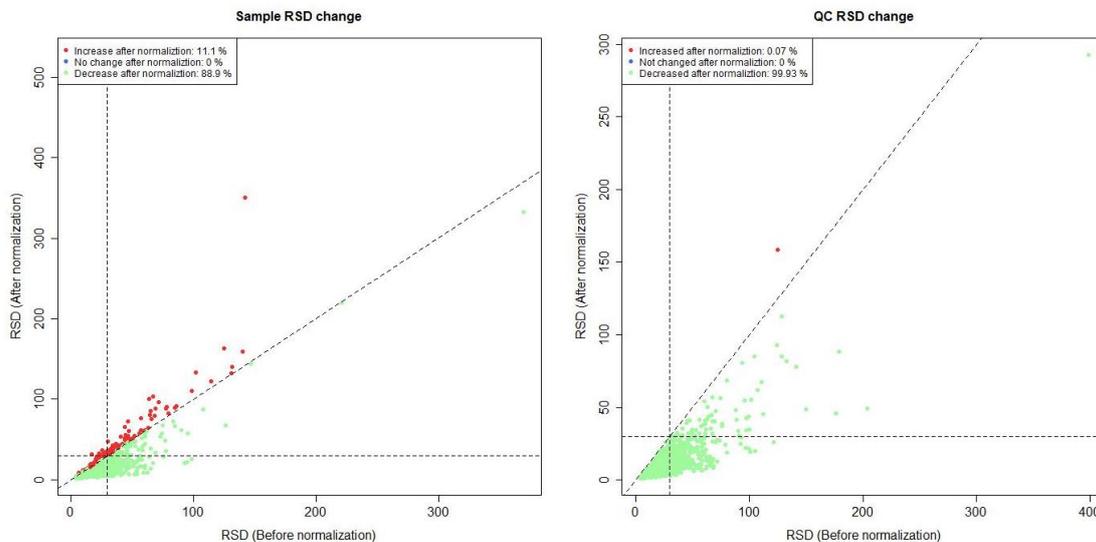
a) “**peak plot**” folder contains peak plots before and after normalization for each peak. An example for peak plot is given below:



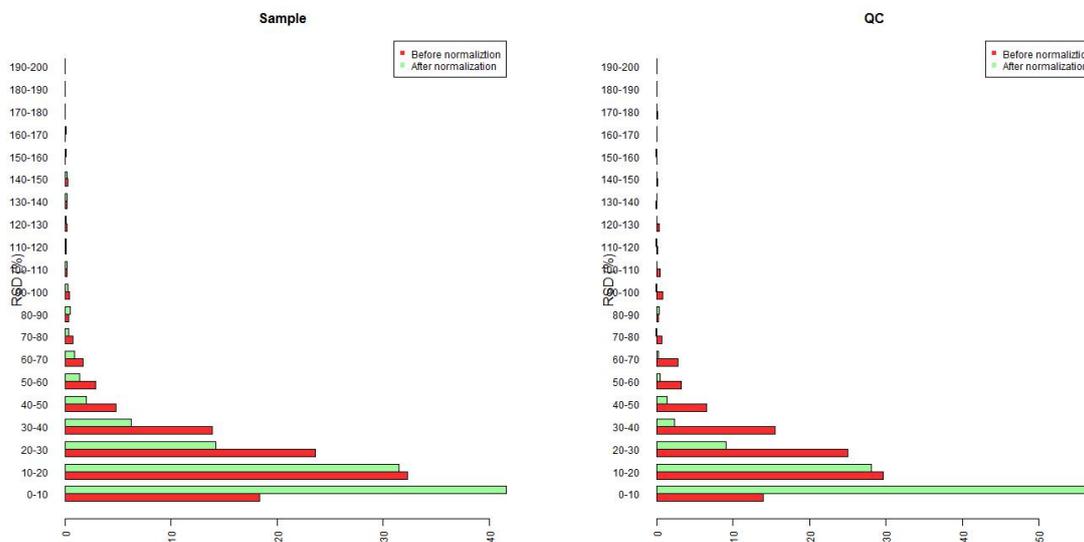
b) “**data svr nor.csv**” is the dataset that contains subject samples, QC samples and sample tags after SVR normalization.

tags			sample and QC RSD		sample and QC data	
feature name	mz	rt	sample.nor.rsd	QC.nor.rsd	S20	S30
1	76.07598	517.004	9.898376	10.01136	1.029944	1.13991
2	83.06036	473.104	4.698322	3.075415	1.010772	1.137051
3	85.02796	435.884	5.253887	3.407347	1.068966	1.175565
4	86.07103	240.986	4.347418	2.060847	1.071258	1.039731
5	104.0701	417.792	8.895212	8.559883	1.040101	1.304325

c) “RSD compare plot.jpeg” is the summarized comparison of RSDs for all peaks before and after normalization. An example for “RSD compare plot.jpeg” is given below:



d) “RSD distribution.jpeg” is the summarized comparison of RSD distributions before and after SVR normalization. An example for “RSD distribution.jpeg” is given below:



- e) “**rsd.change.csv**” is a table that contains the RSDs of all peaks before and after SVR normalization. An example for “rsd.change.csv” is given below:

feature name	mz	rt	sample.rsd	sample.nor.rsd	QC.rsd	QC.nor.rsd
1	76.07598	517.004	15.4572984	9.898376309	19.3035007	10.01135982
2	83.06036	473.104	28.4953914	4.698321856	30.4124249	3.075415173
3	85.02796	435.884	8.085898988	5.253887214	8.55128802	3.407347003
4	86.07103	240.986	7.48780097	4.347418234	7.32711432	2.06084696
5	104.0701	417.792	19.27122234	8.895212435	21.1996386	8.559882973

- f) “normalization file” is an intermediate file during data processing.