User manual

polyPK: An R package for pharmacokinetic analysis of multi-component drugs using a metabolomics approach

Mengci Li, Shouli Wang, Guoxiang Xie, Xiaohui Ma, Tianlu Chen, Wei Jia

Contents

1. Introduction

The advent of mass spectrometry based analytical technologies coupled with multivariate statistical methods offers tremendous new opportunities for understanding the pharmacokinetics of multicomponent herbal medicines. In recent years, Jia's group proposed a Poly-PK strategy, in a few review papers (Jia, et al., 2015; Lan and Jia, 2010; Lan, et al., 2013), to characterize the concentration-time profile and the metabolic response profile of multicomponent drugs, using an integrated phytochemical and metabolomics approach. The brand-new strategy has been successfully applied in examining the complex pharmacokinetics and pharmacodynamics profiles of tea (Xie, et al., 2012) and Huangqi decoction (Xie, et al., 2017). This package is the first implementation of the data analysis steps of Poly-PK strategy with 10 easy-to-use functions. The absorbed drug constituents, the downstream metabolites of the drug constituents and the endogenous metabolites impacted by them can be identified. The association of different types of compounds and their alterations along time can be illustrated by various figures and tables. The representative PK parameters of all compounds can be calculated.

2. Function list

10 functions of the package of polyPK:

DataPre, ***Simi***, ***GetDiffData***, ***GetEndo***, ***GetAbso***, ***GetSecdAbso***, ***PKs***, ***ScatPlot***, ***HeatMap***, ***CorrPlot***

3. Example of a complete analysis

- polyPK::**DataPre**(tes=postData,mv="mean",rz=80,multiple=0.1,sv=TRUE,log=FALSE,filepath=**getwd**())//preprocessing of data
- simi<-polyPK::**Simi**(data1<-preData,data2<-drugData,filepath=**getwd**());

simidata<-simi[[3]]

//get the same component between pre-dose metabolites and drug components

- polyPK::**GetDiffData**(preData, postData,simidata,mv="mean",rz=80,multiple=0.1,sv=TRUE,log=FALSE,t="T test",r.adj="fdr",filepath=**getwd**(),design= FALSE)

// get differential metabolites

- polyPK::**GetEndo**(preData,A,simidata,sim=80,filepath=**getwd**(),design= FALSE)
- polyPK::**GetAbso**(drugData, A, simidata,sim = 80, filepath=**getwd**(),design = FALSE)
- polyPK::**GetSecdAbso**(A,B,C,simidata,sim=80,filepath=**getwd**(),design = FALSE)

// get three groups of the differential metabolites

- polyPK::**PKs**(A,d.point="mean",d.ebar="SE",filepath=**getwd**(),design= FALSE)

// PK analysis

- polyPK::**CorrPlot**(dataset1=B,dataset2=C,cor.method="pearson",filepath=**getwd**(),fig.form="heatmap",design = FALSE) // correlation analysis
- polyPK::**ScatPlot**(scat.data=A,scform="PCA",num.of.cp=2,filepath=**getwd**(),

design = FALSE) // classification analysis

- polyPK::**HeatMap**(data=A,cluster="both",scale="row",filepath=**getwd**(),design= FALSE) // cluster analysis

4. Installation

Input “install.packages(“polyPK”)” in an R console and all the required packages and functions will be installed automatically into your workspace. You can call the functions directly after this.

5. Functions, results and application guidance

DataPre

Preprocess the input data. Variables with a lot of zeros and outliers may be removed. Missing values may be imputed and filled by various methods. Data may be transformed by logarithm transformation.

Arguments

tes — The data under pretreatment (data frame with required format). The first row should be column names. The first and the second columns of the first row should be "Name" and "ID", and you can set 2 more tags at the third and the fourth columns of the first row, such as "m.z" and "RT.min." or anything you like. From the fifth column till the end, sample indexes or names are expected. The first row of the data frame should be the gender information."1"means male, and "2" means female. The second row of the data frame should be the group information. The first column of the second row should be "group", and you can add group indexes of the data from the fifth column at the second row. The format of group number should be "0"(pre-dose). "1","2","3","4"...(post-dose). The third row of the data frame should be the information of time points. Please see the **Fig.S1, Fig.S2** for detailed format.

rz — The percentage of zeros for variable elimination (Default:80). Variables with zero numbers higher than rz.

mv — The method of missing values imputation (Default: "min"). mv=c ("min", "knn", "qrilc"). "min" is the most common method for imputing missing values, which replace the missing values by minimum value in all data sets. "knn" is a complicated method for imputing missing expression data, using nearest neighbor averaging (Troyanskaya, et al., 2001) and all the imputed values are different. Quantile regression approach for left-censored missing (QRILC) imputes missing data by using quantile regression (Lazar, et al., 2016). For dealing with the left-censored missing values in metabolome data, "qrilc" method is the most suitable option (Wei, et al., 2017).

multiple — The special parameter for missing values imputation by “min”. Missing values will be replaced by multiple*min (Default:0.1).

sv — A logical value indicating whether to remove the outliers (Default: TRUE). The data which distance to the mean is bigger than 1.5 times of the difference value between lower quartile and upper quartile, should be identified as an outlier. And it will be replaced by the mean value of the corresponding row.

log — A logical value indicating whether to take the logarithm on the data (Default: FALSE).

filepath — A character string indicating the path where the results may be saved in.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Name	ID	m/z	RT(min)	0-1	0-2	0-3	0-4	0-5	0-6	0-7	0-8	0-9	0-10
2	gender				1	1	1	1	1	2	2	2	2	2
3	group				0	0	0	0	0	0	0	0	0	0
4	timepoints(h)				0	0	0	0	0	0	0	0	0	0
5	octanoyl-rac-glycerol	inx0001	219.1401	3.5638	4.18E+03	5.08E+03	4.77E+03	4.08E+03	2.74E+03	4.99E+03	4.43E+03	4.15E+03	3.54E+03	3.31E+03
6	1,2,3-propanetricarboxylic acid	HMDB31193	73	15.657182	1.10E+07	1.01E+07	1.09E+07	1.45E+07	1.38E+07	1.32E+07	1.18E+07	1.23E+07	1.04E+07	1.34E+07
7	myristic acid	HMDB00806	227.2009	6.8607667	5.78E+05	5.70E+05	4.25E+05	2.89E+05	4.08E+05	6.03E+05	1.68E+05	3.46E+05	4.42E+05	5.87E+05
8	glycolithocholic acid	HMDB00698	432.3107	3.4209	9.71E+02	9.71E+02	9.71E+02	9.71E+02	5.09E+00	9.71E+02	9.71E+02	9.71E+02	9.71E+02	9.71E+02
9	3-indolelactic acid	HMDB00671	203	21.2330824	3.71E+04	6.92E+04	1.23E+05	6.92E+04	9.88E+04	6.11E+04	4.47E+04	6.92E+04	1.00E+05	1.95E+04
10	fructose	HMDB00660	103	16.3498505	7.89E+05	4.71E+05	4.12E+05	5.20E+05	3.27E+05	6.03E+05	4.32E+05	5.58E+05	4.69E+05	6.37E+05
11	phenylpyruvic acid	HMDB00205	165.0545	3.5218833	1.26E+04	1.36E+04	2.01E+04	1.91E+04	1.93E+04	1.63E+04	1.95E+04	2.10E+04	1.73E+04	2.19E+04
12	heptadecanoic acid	HMDB02259	269.2477	9.5559	8.97E+04	8.97E+04	8.68E+04	4.28E+04	7.08E+04	6.84E+04	2.81E+04	3.46E+04	7.21E+04	6.52E+04
13	phosphoric acid	HMDB02142	98.9846	4.3821	4.44E+03	3.69E+03	3.22E+03	4.32E+03	4.07E+03	4.44E+03	5.01E+03	4.29E+03	2.84E+03	3.33E+03

Figure S1. The format of pre-dose metabolome data.

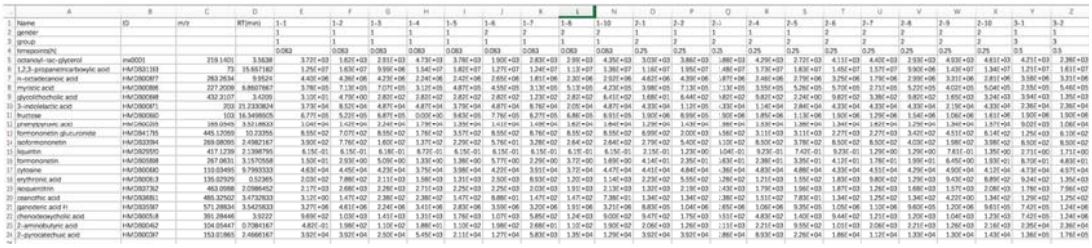

Figure S2. The format of post-dose metabolome data.

Usage and results

```
library(polyPK)

data("postData")
pred_post<-polyPK::DataPre(tes=postData,mv="min",rz=80,sv=TRUE,log=FALSE,filepath=getwd())
```

An example of preprocessed post-dose dataset:

```
pred_post[c(1:10),c(1:10)]
```

```
##                                    Name          ID       m.z    RT.min.      X1.1      X1.2      X1.3      X1.4      X1.5      X1.6
## 1                                gender        <NA>       <NA>         1         1         1         1         1         2
## 2                                 group        <NA>       <NA>         1         1         1         1         1         1
## 3                          timepoints(h)        <NA>       <NA>     0.083     0.083     0.083     0.083     0.083     0.083
## 4                 octanoyl-rac-glycerol  inx0001 219.1401  3.5638000 3.72e+03  1.62e+03 2.91e+03 4.73e+03 3.78e+03 1.90e+03
## 5  1,2,3-propanetricarboxylic acid HMDB31193   73.0000 15.6571820 1.25e+07  1.63e+07 9.99e+06 1.54e+07 1.82e+07 1.27e+07
## 6                              tryptophan HMDB00929 203.0819   1.3524000 7.24e+04  9.80e+04 1.14e+05 9.10e+04 9.62e+04 6.97e+04
## 7                  n-octadecanoic acid HMDB00827 283.2634   9.9524000 4.40e+06  4.36e+06 4.23e+06 2.24e+06 2.42e+06 2.65e+06
## 8                        myristic acid HMDB00806 227.2009   6.8607667 3.78e+05  7.13e+05 7.07e+05 3.12e+05 4.87e+05 4.55e+05
## 9               glycolithocholic acid  HMDB00698 432.3107   3.4209000 3.10e+01  4.79e+00 2.82e+02 2.82e+02 2.82e+02 2.82e+02
## 10                3-indolelactic acid HMDB00671 203.0000 21.2330824 3.73e+04  8.52e+04 4.87e+04 4.87e+04 3.79e+04 4.87e+04
```

Simi

A function which can get the similar metabolites of two data. Especially the similar metabolites between drug and pre-dose metabolites.

	A	B	C	D	E	F	G
1	Name	ID	m/z	RT(min)	herbal-1	herbal-2	herbal-3
2	formonoetin	HMDB05808	267.0631	3.1571	4752.02	5569.88	4380.11
3	glabridin	HMDB34188	323.1302	3.8649	2.88	4.96	3.13
4	glycyrrhetinic acid	HMDB11628	469.3303	4.8802	20.27	0.00	0.00
5	isoquercitrin	HMDB37362	463.0988	2.0986	4.14	6.44	5.53
6	kaempferol	HMDB05801	285.0491	2.8928	21.02	12.31	18.33
7	L-canaline	HMDB12251	135.0764	0.6769	51112.77	44132.60	40890.61
8	liquiritin	HMDB29520	417.1239	2.1399	34289.13	30909.40	31102.67
9	pentoxifylline	HMDB14944	279.1450	0.5721	3877.70	2871.79	2844.60
10	sinensal	HMDB13693	219.1702	4.4133	6397.82	7676.32	7809.47
11	p-coumaric acid	HMDB30677	163.04003	1.373283	2377.70	2651.84	2198.19

Figure S3. The format of drug constituents' data.

Arguments

data1 The pre-dose dataset (data frame with required format). The first row should be column names. The first and the second columns of the first row should be "Name" and "ID", and you can set 2 more tags at the third and the fourth columns of the first row, such as "m.z" and "RT.min." or anything you like. From the fifth column till the end, sample indexes or names are expected. The first row of the data frame should be the gender information."1"means male, and "2" means female. The second row of the data frame should be the group information. The first column of the second row should be "group", and you can add group indexes of the data from the fifth column at the second row. The format of group number should be "0"(pre-dose). "1","2","3","4"...(post-dose). The third row of the data frame should be the information of time points. Please see **Fig.S1** for detailed format.

data2 The drug constituents dataset (data frame), see **Fig.S3**.

filepath A character string indicating the path where the results may be saved in.

Usage and results

```
data("preData")
data("drugData")
Simi(data1<-preData,data2<-drugData,filepath=getwd())

## $`repetitive rates in data1`
## [1] 0.3076923
##
## $`repetitive rates in data2`
## [1] 0.6666667
##
## $`similar metabolites`
##       data1[-c(1:3), 2]
## [1,] "HMDB00205"
## [2,] "HMDB00806"
## [3,] "HMDB00827"
## [4,] "inx0001"
```

GetDiffData

A function to get all the differential compounds between the pre-dose and every post-dose datasets

Arguments

preData The original pre-dose dataset (data frame) with an indicator of gender variable at the first row, grouping variable at the second row, and time points at the third row. Please see the **Fig.S1** for detailed format.

postData The original post-dose dataset (data frame) with an indicator of gender variable at the first row, grouping variable at the second row, and time points at the third row. Please see the **Fig.S2** for detailed format.

simidata The same compounds of drug and pre-dose metabolome data, which are derived from Simi.

rz The percentage of zeros for variable elimination(Default:80)

mv The method of missing values imputation (Default: "mean"). mv=c("min","knn","qrilc")

sv A logical value indicating whether to remove the outliers (Default: TRUE). The data which distance to the mean is bigger than 1.5 times of the difference value between lower quartile and upper quartile, should be identified as an outlier. And it will be replaced by the mean value of the corresponding row.

log A logical value indicating whether to take the logarithm on the datasets (Default: FALSE)

t The method for differential compounds identification. C ("Ttest", "MWtest"). Default: "Ttest". Compounds with p values less than 0.05 were taken as differential ones. We suggest the users that choosing "MWtest" (Mann–Whitney test) is better. Unlike the t-test, Mann–Whitney test does not require the assumption of normal distributions (Khodakarim, et al., 2014).

r.adj The methods for p values adjustment. r.adj=c("holm", "fdr"). Default: "fdr". "holm" is intended to control the familywise error rate and offers a simple test uniformly more powerful than the Bonferroni correction (Holm, 1979). FDR-controlling procedures provide less stringent control of Type I errors compared to familywise error rate (FWER) controlling procedures (such as the "holm" correction), which control the probability of at least one Type I error. Thus, FDR-controlling procedures have greater power, at the cost of increased numbers of Type I errors (Benjamini, et al., 2001). The "fdr" method is less strict than "holm".

filepath A character string indicating the path where the results may be saved in.

design (optional) a study design dataset (data frame with required format). Default: "FALSE", see **Fig.S4** for detailed.

	A	B	C	D	E	F	G	H
1	1.Basic information		local time					
2	Study number	2012ZX001						
3	Study title and aim	PKs of Huangqi decoction						
4	Institution and operator	Jia's lab, Shouli Wang						
5	Experiment date	2017/5/8~2017/5/10						
6								
7	2.Subject information							
8	Total subject number	10						
9	Subject type	Human						
10	Subject species	Plasma						
11								
12	3.Drug information							
13	Drug name	Huangqi Decoction						
14	Main constitutes	Huangqi;Gancao						
15	Drug origin/producer	Jiangyin						
16	Dose form	Granules						
17	Drug part	Root						
18								
19	4.Study information							
20	Number of pre-dose timepoints	1	0					
21	Number of post-dose timepoints(h)	14	0.083	0.25	0.5	1	2	4
22	Number of meals(h)	4	4.5	10.5	28.5	34.5		
23	Number of sleep(h)	2	13	37				
24	Sample preparation	NA						
25	Sample analysis	GC-TOF/MS;UPLC-QTOF/MS						
26	Data processing	MarkerLynx;QI						
27	Standardized meal?	Yes						
28	Drug dosage	140g/person						
29	Drug administration time	1	8:00					

Figure S4. The format of study design (basic information).

Usage and results

```
data("preData")
data("postData")
data("design")
data("simidata")
dif<-GetDiffData(preData,postData,simidata,mv="min",rz=80,sv=TRUE,log=FALSE,t="T
test",r.adj="fdr",filepath=getwd(),design=design)
```

Figure S5. The results in file folder.

The results differential metabolites and p-values are listed in weight rank order which was calculated by the SAM (Significance analysis of microarrays) method (Tusher, et al., 2001). SAM is a method for identifying genes on a microarray with statistically significant changes in expression. Here, the function assigns a score to each metabolite, uses permutations to estimate the percentage of metabolites identified by chance, and relists the metabolites based on the scores.

An example of the differential compounds, with preprocessed data:

```
prepoA<-dif$A
as.data.frame(prepoA[,c(1:12)])
```

```
##                                    Name           ID        m.z      RT.min.       X0.1
## 1                                gender         <NA>        <NA>          <NA>          1
## 2                                 group         <NA>        <NA>          <NA>          0
## 3                         timepoints(h)         <NA>        <NA>          <NA>          0
## 110          glycolithocholic acid HMDB00698 432.3107    3.420900        971
## 210                          fructose HMDB00660 103.0000   16.349850     789000
## 32                     myristic acid HMDB00806 227.2009    6.860767     578000
## 4                n-octadecanoic acid HMDB00827 283.2634    9.952400    3250000
## 5                        tryptophan HMDB00929 203.0819    1.352400      76200
## 6                phenylpyruvic acid HMDB00205 165.0545    3.521883      12600
## 7             octanoyl-rac-glycerol     inx0001 219.1401    3.563800       4180
## 8  1,2,3-propanetricarboxylic acid HMDB31193  73.0000   15.657182 11000000
## 9                3-indolelactic acid HMDB00671 203.0000   21.233082      37100
## 10         formononetin glucuronide HMDB41735 445.1206  10.2335500       <NA>
## 11                    gancaonin V HMDB37586 311.1301   2.2231667       <NA>
## 12                isoquercitrin HMDB37362 463.0988   2.0986452       <NA>
## 13                ceanothic acid HMDB36851 485.3250   3.4732833       <NA>
## 14               ganoderic acid H HMDB35987 571.2883   3.5425833       <NA>
## 15                isoformononetin HMDB33994 269.0809   2.4982167       <NA>
## 16                     liquiritin HMDB29520 417.1239   2.1398795       <NA>
## 17                  formononetin HMDB05808 267.0631   3.1570558       <NA>
## 18                  trigonelline HMDB00875 136.0404   2.5085333       <NA>
## 19                     histamine HMDB00870 112.0763   0.7084167       <NA>
## 20                phytanic acid HMDB00801 311.2948  10.8229833       <NA>
## 21            cholesterol sulfate HMDB00653 465.3041   6.6551167       <NA>
## 22                     cytosine HMDB00630 110.0349   9.7993333       <NA>
## 23               erythronic acid HMDB00613 135.0293   0.5236500       <NA>
## 24          chenodeoxycholic acid HMDB00518 391.2845   3.9222000       <NA>
```

```
## 25                2-aminobutyric acid HMDB00452 104.0545   0.7084167       <NA>
## 26              2-pyrocatechuic acid HMDB00397 153.0187   2.4666167       <NA>
## 27                              urea HMDB00294 100.0000   7.1697798       <NA>
## 28                   octadecanamide HMDB34146 284.2946    8.871300     119000
## 29                heptadecanoic acid HMDB02259 269.2477    9.555900      89700
## 30                   phosphoric acid HMDB02142  98.9846    4.382100       4440
## 31                hyodeoxycholic acid HMDB00733 391.2842    3.574000      33500
##          X0.2     X0.3     X0.4     X0.5     X0.6     X0.7     X0.8
## 1           1        1        1        1        2        2        2
## 2           0        0        0        0        0        0        0
## 3           0        0        0        0        0        0        0
## 110       971      971      971 5.09e+00      971      971      971
## 210    471000   412000   520000 3.27e+05   603000   432000   558000
## 32     570000   425000   289000 4.08e+05   603000   168000   346000
## 4     2860000  3110000  2160000 2.79e+06  3250000  1610000  2030000
## 5      101000   125000    96100 1.18e+05    87900   101000   115000
## 6       13600    20100    19100 1.93e+04    16300    19500    21000
## 7        5080     4770     4080 2.74e+03     4990     4430     4150
## 8    10100000 10900000 14500000 1.38e+07 13200000 11800000 12300000
## 9       69200   123000    69200 9.88e+04    61100    44700    69200
## 10       <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
## 11       <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
## 12       <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
## 13       <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
## 14       <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
## 15       <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
## 16       <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
## 17       <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
## 18       <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
## 19       <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
## 20       <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
## 21       <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
## 22       <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
## 23       <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
## 24       <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
## 25       <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
## 26       <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
## 27       <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
## 28      99100   159000   264000 9.54e+04   361000   361000   267000
## 29      89700    86800    42800 7.08e+04    68400    28100    34600
## 30       3690     3220     4320 4.07e+03     4440     5010     4290
## 31       9930    22200    13300 2.57e+04     7800    24600    33500
```

An example of the differential compounds, with original data:

```
orgA<-dif$A_pre
as.data.frame(orgA[,c(1:12)])
```

```
##                                    Name            ID         m.z      RT.min.      X0.1
## 1                                gender          <NA>         <NA>         <NA>         1
## 2                                 group          <NA>         <NA>         <NA>         0
## 3                       timepoints(h)            <NA>         <NA>         <NA>         0
## 110          glycolithocholic acid  HMDB00698  432.3107       3.4209          971
## 210                        fructose HMDB00660          103 16.3498505   789000
## 32                  myristic acid HMDB00806   227.2009   6.8607667   578000
## 4                n-octadecanoic acid HMDB00827   283.2634       9.9524 3250000
## 5                       tryptophan HMDB00929   203.0819       1.3524    76200
## 6                 phenylpyruvic acid HMDB00205   165.0545   3.5218833    12600
## 7              octanoyl-rac-glycerol     inx0001   219.1401       3.5638      4180
## 8    1,2,3-propanetricarboxylic acid HMDB31193          73   15.657182 1.1e+07
## 9                3-indolelactic acid HMDB00671          203 21.2330824    37100
## 10          formononetin glucuronide HMDB41735 445.12059     10.23355      <NA>
## 11                  gancaonin V HMDB37586 311.13009   2.2231667      <NA>
## 12                isoquercitrin HMDB37362   463.0988   2.0986452      <NA>
## 13                ceanothic acid HMDB36851 485.32502   3.4732833      <NA>
## 14              ganoderic acid H HMDB35987 571.28834   3.5425833      <NA>
## 15               isoformononetin HMDB33994 269.08095   2.4982167      <NA>
## 16                    liquiritin HMDB29520   417.1239   2.1398795      <NA>
## 17                formononetin HMDB05808   267.0631   3.1570558      <NA>
## 18                trigonelline HMDB00875 136.04045   2.5085333      <NA>
## 19                   histamine HMDB00870 112.07626   0.7084167      <NA>
## 20                phytanic acid HMDB00801 311.29478 10.8229833      <NA>
## 21            cholesterol sulfate HMDB00653 465.30412   6.6551167      <NA>
## 22                   cytosine HMDB00630 110.03495   9.7993333      <NA>
## 23              erythronic acid HMDB00613 135.02929      0.52365      <NA>
## 24           chenodeoxycholic acid HMDB00518 391.28446       3.9222      <NA>
## 25             2-aminobutyric acid HMDB00452 104.05447   0.7084167      <NA>
```

```
## 26                2-pyrocatechuic acid HMDB00397 153.01865   2.4666167      <NA>
## 27                                urea HMDB00294         100   7.1697798      <NA>
## 28                 octadecanamide HMDB34146   284.2946       8.8713   119000
## 29               heptadecanoic acid HMDB02259   269.2477       9.5559    89700
## 30                  phosphoric acid HMDB02142    98.9846       4.3821     4440
## 31               hyodeoxycholic acid HMDB00733   391.2842        3.574    33500
##          X0.2     X0.3     X0.4     X0.5     X0.6     X0.7     X0.8
## 1           1        1        1        1        2        2        2
## 2           0        0        0        0        0        0        0
## 3           0        0        0        0        0        0        0
## 110       971      971      971     5.09      971      971      971
## 210    471000   412000   520000   327000   603000   432000   558000
## 32     570000   425000   289000   408000   603000   168000   346000
## 4     2860000  3110000  2160000  2790000  3250000  1610000  2030000
## 5      101000   125000    96100   118000    87900   101000   115000
## 6       13600    20100    19100    19300    16300    19500    21000
## 7        5080     4770     4080     2740     4990     4430     4150
## 8    10100000 10900000 14500000 13800000 13200000 11800000 12300000
## 9       69200   123000    69200    98800    61100    44700    69200
……

## 28      99100   159000   264000    95400   361000   361000   267000
## 29      89700    86800    42800    70800    68400    28100    34600
## 30       3690     3220     4320     4070     4440     5010     4290
## 31       9930    22200    13300    25700     7800    24600    33500
```

The p values:

```
p<-dif$p
p
```

```
##                                          [,1]          [,2]          [,3]
## glycolithocholic acid          0.0001194918 8.082684e-01 1.790408e-05
## fructose                       0.0984948479 1.007318e-05 7.965273e-07
## myristic acid                  0.5815884996 4.024950e-01 4.600790e-01
## n-octadecanoic acid            0.3074598184 1.213953e-01 7.680574e-01
## tryptophan                     0.0202897340 3.684567e-02 2.457435e-03
## phenylpyruvic acid             0.0251127544 1.178979e-02 2.273702e-03
## octanoyl-rac-glycerol          0.0260244164 4.654868e-01 5.919837e-02
## 1,2,3-propanetricarboxylic acid 0.1460157496 2.074242e-02 1.922077e-01
## 3-indolelactic acid            0.1123450519 6.350310e-02 1.309725e-03
##                                          [,4]          [,5]
## glycolithocholic acid          7.473468e-02 7.119534e-01
## fructose                       1.293703e-05 3.866770e-02
## myristic acid                  5.494115e-03 1.007495e-01
## n-octadecanoic acid            1.369011e-02 1.061509e-01
## tryptophan                     1.884732e-03 4.076133e-05
## phenylpyruvic acid             8.147760e-04 2.044026e-04
## octanoyl-rac-glycerol          1.721622e-02 6.383351e-05
## 1,2,3-propanetricarboxylic acid 6.119316e-02 7.879574e-02
## 3-indolelactic acid            4.042092e-04 2.215191e-03
```

The adjusted p values:

```
padj<-dif$p_adj
padj
```

```
##                                          [,1]          [,2]          [,3]
## glycolithocholic acid          0.0007681614 0.8082683735 2.014209e-04
## fructose                       0.1416789827 0.0001940554 3.584373e-05
## myristic acid                  0.6231305353 0.4644173386 5.109002e-01
## n-octadecanoic acid            0.3640971533 0.1560796964 7.855133e-01
## tryptophan                     0.0444480340 0.0690856278 7.372306e-03
## phenylpyruvic acid             0.0509173364 0.0312082673 7.308328e-03
## octanoyl-rac-glycerol          0.0509173364 0.5109001872 1.019886e-01
## 1,2,3-propanetricarboxylic acid 0.1825196870 0.0444480340 2.337661e-01
## 3-indolelactic acid            0.1486919805 0.1020585555 5.357964e-03
##                                          [,4]          [,5]
## glycolithocholic acid          0.1159676132 0.7450674653
```

```
## fructose                                0.0001940554  0.0696018547
## myristic acid                           0.0154521974  0.1416789827
## n-octadecanoic acid                     0.0342252873  0.1447511932
## tryptophan                              0.0070677432  0.0003668520
## phenylpyruvic acid                      0.0036664920  0.0011497647
## octanoyl-rac-glycerol                   0.0407752660  0.0004787513
## 1,2,3-propanetricarboxylic acid 0.1019886063  0.1181936146
## 3-indolelactic acid                     0.0020210459  0.0073083285
```

GetEndo

A function to get the altered endogenous metabolites by similarity analysis on the list of differential compounds and the list of pre-dose compounds.

Arguments

pre	The pre-dose dataset (data frame).
A	The differential compounds which are derived from the GetDiffData function.
simidata	The same compounds of drug and pre-dose metabolome data, which are derived from Simi.
sim	The parameter (percentage) for similarity analysis. Default: 80.
filepath	A character string indicating the path where the results may be saved in.
design	(optional) a study design dataset (data frame with required format). See **Fig.S4** for detailed format. Default: "FALSE"

Usage and results

```
 data("preData")
 data("A")
 data("design")
 data("simidata")
endo<-GetEndo(preData,A,simidata,sim=80,filepath=getwd(),design=design)

endo[,c(1:12)]
```

```
##                                      Name           ID         m.z     RT.min.        X0.1
## 1                                  gender                    <NA>        <NA>           1
## 2                                   group                    <NA>        <NA>           0
## 3                          timepoints(h)                     <NA>        <NA>           0
## 7                  octanoyl-rac-glycerol     inx0001  219.1401   3.563800        4180
## 28                       octadecanamide HMDB34146  284.2946   8.871300      119000
## 8    1,2,3-propanetricarboxylic acid HMDB31193   73.0000 15.657182 11000000
## 29                   heptadecanoic acid HMDB02259  269.2477   9.555900       89700
## 30                       phosphoric acid HMDB02142   98.9846   4.382100        4440
## 5                          tryptophan HMDB00929  203.0819   1.352400       76200
## 4                    n-octadecanoic acid HMDB00827  283.2634   9.952400     3250000
## 32                         myristic acid HMDB00806  227.2009   6.860767      578000
## 31                   hyodeoxycholic acid HMDB00733  391.2842   3.574000       33500
## 110                glycolithocholic acid   HMDB00698  432.3107   3.420900         971
## 9                    3-indolelactic acid HMDB00671  203.0000  21.233082       37100
## 210                          fructose HMDB00660  103.0000  16.349850      789000
## 6                     phenylpyruvic acid HMDB00205  165.0545   3.521883       12600
##           X0.2       X0.3       X0.4       X0.5       X0.6       X0.7       X0.8
## 1            1          1          1          1          2          2          2
## 2            0          0          0          0          0          0          0
## 3            0          0          0          0          0          0          0
## 7         5080       4770       4080   2.74e+03       4990       4430       4150
## 28       99100     159000     264000   9.54e+04     361000     361000     267000
## 8     10100000   10900000   14500000   1.38e+07   13200000   11800000   12300000
```

```
## 29      89700     86800      42800  7.08e+04     68400     28100     34600
## 30       3690      3220       4320  4.07e+03      4440      5010      4290
## 5      101000    125000      96100  1.18e+05     87900    101000    115000
## 4     2860000   3110000    2160000  2.79e+06   3250000   1610000   2030000
## 32     570000    425000     289000  4.08e+05    603000    168000    346000
## 31       9930     22200      13300  2.57e+04      7800     24600     33500
## 110       971       971        971  5.09e+00       971       971       971
## 9       69200    123000      69200  9.88e+04     61100     44700     69200
## 210    471000    412000     520000  3.27e+05    603000    432000    558000
## 6       13600     20100      19100  1.93e+04     16300     19500     21000
```

GetAbso

A function to get the absorbed drug constituents by similarity analysis on the list of differential compounds and the list of drug constituents:

Arguments

drug	The drug constituents dataset (data frame)
A	The differential compounds which are derived from the GetDiffData function.
simidata	The same compounds of drug and pre-dose metabolome data, which are derived from Simi.
sim	The parameter (percentage) for similarity analysis. Default: 80.
filepath	A character string indicating the path where the results may be saved in.
design	(optional) a study design dataset (data frame with required format). See **Fig.S4** for detailed format. Default: "FALSE"

Usage and results

```
abso<-GetAbso(drugData, A, simidata,sim = 80, filepath=getwd(),design = design)

abso[,c(1:14)]
```

```
##                            Name            ID         m.z       RT.min.       X0.1       X0.2
## 1                        gender          <NA>         <NA>            1          1
## 2                         group          <NA>         <NA>            0          0
## 3                timepoints(h)          <NA>         <NA>            0          0
## 7   octanoyl-rac-glycerol     inx0001  219.1401     3.563800       4180       5080
## 16                liquiritin  HMDB29520  417.1239    2.1398795       <NA>       <NA>
## 17              formononetin  HMDB05808  267.0631    3.1570558       <NA>       <NA>
## 4     n-octadecanoic acid  HMDB00827  283.2634     9.952400    3250000    2860000
## 32            myristic acid  HMDB00806  227.2009     6.860767     578000     570000
## 6       phenylpyruvic acid  HMDB00205  165.0545     3.521883      12600      13600
##           X0.3       X0.4       X0.5       X0.6       X0.7       X0.8       X0.9      X0.10
## 1            1          1          1          2          2          2          2          2
## 2            0          0          0          0          0          0          0          0
## 3            0          0          0          0          0          0          0          0
## 7         4770       4080  2.74e+03       4990       4430       4150       3540       3310
## 16        <NA>       <NA>       <NA>       <NA>       <NA>       <NA>       <NA>       <NA>
## 17        <NA>       <NA>       <NA>       <NA>       <NA>       <NA>       <NA>       <NA>
## 4     3110000    2160000  2.79e+06    3250000    1610000    2030000    2970000    2840000
## 32     425000     289000  4.08e+05     603000     168000     346000     442000     587000
## 6       20100      19100  1.93e+04      16300      19500      21000      17300      21900
```

GetSecdAbso

A function to get secondary metabolites of the absorbed drug constituents:

Arguments

A The differential compounds dataset which is derived from the GetDiffData function.

B The altered endogenous metabolites dataset which is derived from the GetEndo function.

C The absorbed drug constituents' dataset which is derived from the GetAbso function.

simidata The same compounds of drug and pre-dose metabolome data, which are derived from Simi.

sim The parameter (percentage) for similarity analysis. Default: 80.

filepath A character string indicating the path where the results may be saved in.

design (optional) a study design dataset (data frame with required format). See **Fig.S4** for detailed format. Default: "FALSE"

Usage and results

secabso<-GetSecdAbso(A,B,C,simidata,sim=80,filepath=**getwd**(),design)

secabso[,**c**(1:12)]

```
##                                      Name           ID        m.z      RT.min.       X0.1
## 1                                  gender         <NA>        <NA>        <NA>          1
## 2                                   group         <NA>        <NA>        <NA>          0
## 3                          timepoints(h)          <NA>        <NA>        <NA>          0
## 10          formononetin glucuronide HMDB41735  445.1206 10.2335500        <NA>
## 11                      gancaonin V HMDB37586  311.1301   2.2231667        <NA>
## 12                    isoquercitrin HMDB37362  463.0988   2.0986452        <NA>
## 13                   ceanothic acid HMDB36851  485.3250   3.4732833        <NA>
## 14                 ganoderic acid H HMDB35987  571.2883   3.5425833        <NA>
## 15                  isoformononetin HMDB33994  269.0809   2.4982167        <NA>
## 8   1,2,3-propanetricarboxylic acid HMDB31193   73.0000  15.657182 11000000
## 5                       tryptophan HMDB00929  203.0819    1.352400     76200
## 18                    trigonelline HMDB00875  136.0404   2.5085333        <NA>
## 19                       histamine HMDB00870  112.0763   0.7084167        <NA>
## 20                   phytanic acid HMDB00801  311.2948  10.8229833        <NA>
## 31              hyodeoxycholic acid HMDB00733  391.2842    3.574000       33500
## 110          glycolithocholic acid  HMDB00698  432.3107    3.420900         971
## 21             cholesterol sulfate HMDB00653  465.3041   6.6551167        <NA>
## 22                         cytosine HMDB00630  110.0349   9.7993333        <NA>
## 23                erythronic acid HMDB00613  135.0293   0.5236500        <NA>
## 24            chenodeoxycholic acid HMDB00518  391.2845   3.9222000        <NA>
## 25              2-aminobutyric acid HMDB00452  104.0545   0.7084167        <NA>
## 26             2-pyrocatechuic acid HMDB00397  153.0187   2.4666167        <NA>
## 27                             urea HMDB00294  100.0000   7.1697798        <NA>
##          X0.2      X0.3      X0.4      X0.5      X0.6      X0.7      X0.8
## 1           1         1         1         1         2         2         2
## 2           0         0         0         0         0         0         0
## 3           0         0         0         0         0         0         0
……

## 8    10100000  10900000  14500000  1.38e+07  13200000  11800000  12300000
## 5      101000    125000     96100  1.18e+05     87900    101000    115000
……

## 31       9930     22200     13300  2.57e+04      7800     24600     33500
```

PKs

A function to calculate the seven representative PK parameters (maximum plasma concentration (Cmax), the time to reach Cmax (Tmax), area under the

concentration-time curve (AUC), the rate of clear (CL), the last time-point (Tlast), the first time-point (Tfirst) and the least plasma concentration of a drug after administration (Cmin)) and plot the time-intensity curves (**Fig.S6)** for specified compounds.

Arguments

d.pk The data under analysis (data frame with required format). Please see the **Fig.S1, Fig.S2** for detailed format.

d.point The value to calculate the pharmacokinetics parameters, and the value of points in the time-intensity curve. d.point=c ("mean", "median"). Default: "mean".

d.ebar The value of error bars. d.ebar=c ("SE","SD"). Default: "SE".

filepath A character string indicating the path where the results may be saved in.

design (optional) a study design dataset (data frame with required format). See **Fig.S4** for detailed format. Default: "FALSE"

Usage and results

A list of metabolites and 7 pharmacokinetics parameters (Tmax, Cmax, AUC, CL, Tlast, Tfirst, Cmin) of specified compounds:

```
data("B")
data("design")
pks<-PKs(B,d.point="mean",d.ebar="SE",filepath=getwd(),design=design)
knitr::kable(pks[c(1:9),],align = 'c')
```

	Name	ID	m.z	R.T.min.	Tmax	Tlast	Tfirst	Cmax	Cmin	AUC	CL
7	octanoyl-rac-glycerol	inx0001	219.1401	3.563800	0	5	0	4127	2458	17010.01	0.0008818
28	octadecanamide	HMDB34146	284.2946	8.871300	0	0	0	216850	0	108425.00	0.0001383
29	heptadecanoic acid	HMDB02259	269.2477	9.555900	0	0	0	64820	0	32410.00	0.0004628
30	phosphoric acid	HMDB02142	98.9846	4.382100	0	0	0	3965	0	1982.50	0.0075662
4	n-octadecanoic acid	HMDB00827	283.2634	9.952400	2	5	0	3206000	2065000	13553175.68	0.0000011
32	myristic acid	HMDB00806	227.2009	6.860767	2	5	0	497400	268300	2027495.56	0.0000074
9	3-indolelactic acid	HMDB00671	203.0000	21.233082	5	5	0	110970	15000	218756.92	0.0000686
210	fructose	HMDB00660	103.0000	16.349850	3	5	0	1660600	401000	5447994.68	0.0000028
6	phenylpyruvic acid	HMDB00205	165.0545	3.521883	0	5	0	18070	12380	71589.80	0.0002095

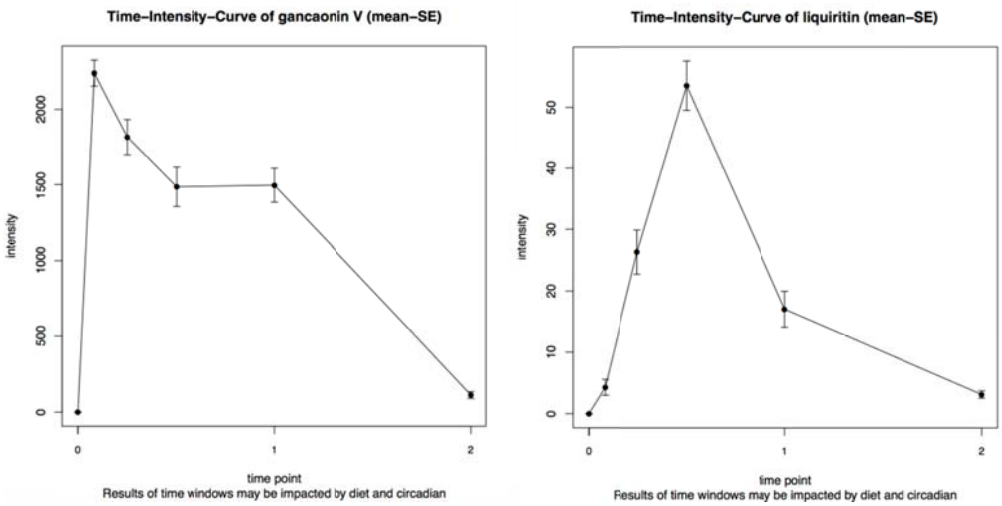

Figure S6. The example time-intensity curves of gancaonin and liquiritin.

The time-intensity curves of specified compounds are in the folder named “PKs” which is created automatically.

CorrPlot

A function to calculate the correlation coefficients and plot the correlation diagrams (8 types) of two input datasets.

Arguments

dataset1	The first dataset (data frame with required format). Please see **Fig.S1**, **Fig.S2** for detailed format. This variable maybe the results of GetEndo,GetAbso,GetSecdAbso.
dataset2	The second dataset (data frame with required format). The form of dataset2 is the same as the form of dataset1. This variable maybe the results of GetEndo,GetAbso,GetSecdAbso.
cor.method	A character string indicating which correlation analysis ("pearson", "kendall", or "spearman") is to be used. Default: "spearman".
filepath	A character string indicating the path where the results may be saved in.
fig.form	The form of the correlation diagram. figure.fig.form=c("heatmap","bubble","ordered.bubble","chord","square","ord.square","pie","ord.pie"). Default: "heatmap".
design	(optional) a study design dataset (data frame with required format). See **Fig.S4** for detailed format. Default: "FALSE"

Usage and results

```
data("B")
data("C")
```

CorrPlot(dataset1=B,dataset2=C,cor.method="pearson",filepath=**getwd**(),fig.form="heat map",design = design)

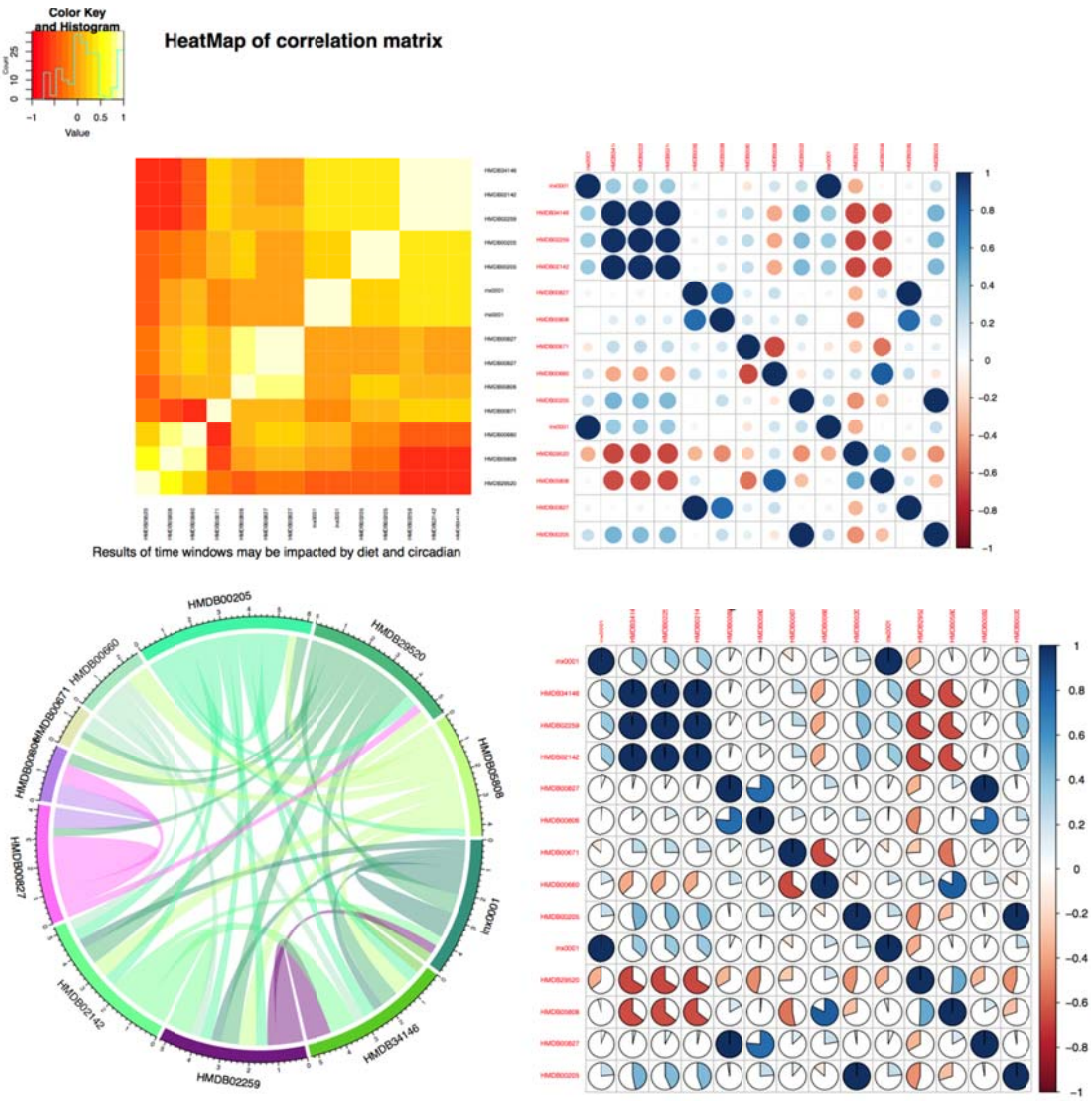

Figure S7. The results of the "heatmap", "bubble", "chord" and "pie" forms of correlation diagram.

ScatPlot

A function to plot the PCA or PLSDA score figures and trajectories on input data

Arguments

scat.data	The data under analysis (data frame with required format). Please see **Fig.S1**, **Fig.S2** for detailed format.

scform	The form of scat plot. scform=c ("PCA","PLSDA"). Default: "PCA".
num.of.cp	The number of components to decompose. Default:2.
fold	Integer: number of random permutations [default is 100 for single response models]
filepath	A character string indicating the path where the results may be saved in.
design	(optional) a study design dataset (data frame with required format). See **Fig.S4** for detailed format. Default: "FALSE"

Usage and results

PLS-DA produces a matrix of classification error rate estimation (**Fig.S9**). The dimensions correspond to the components in the model and to the prediction method used, respectively.

```
data("A")
ScatPlot(scat.data=A,scform="PLSDA",num.of.cp=2,filepath=getwd(),design)
```

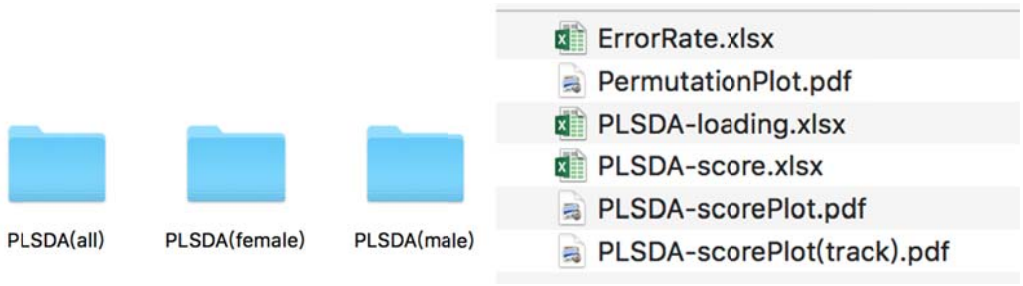

Figure S8. The results folders and files of PLS-DA.

	A	B	C	D
1		max.dist	centroids.dist	mahalanobis.dist
2	comp 1	0.68	0.58	0.58
3	comp 2	0.52	0.37	0.42

Figure S9. The error rates of PLS-DA.

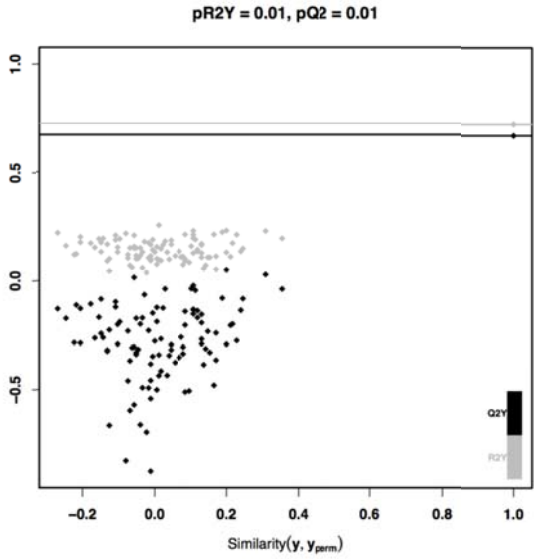

Figure S10. The R2-Q2 scatter plot of permutation (n=100) on PLS-DA.

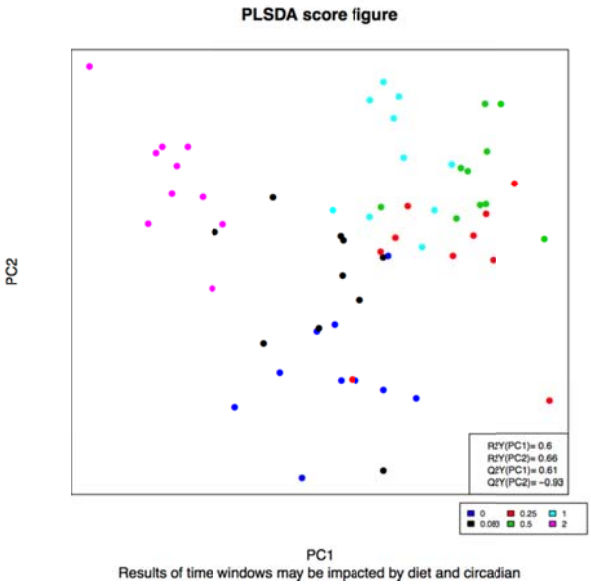

Figure S11. The score plot of PLS-DA.

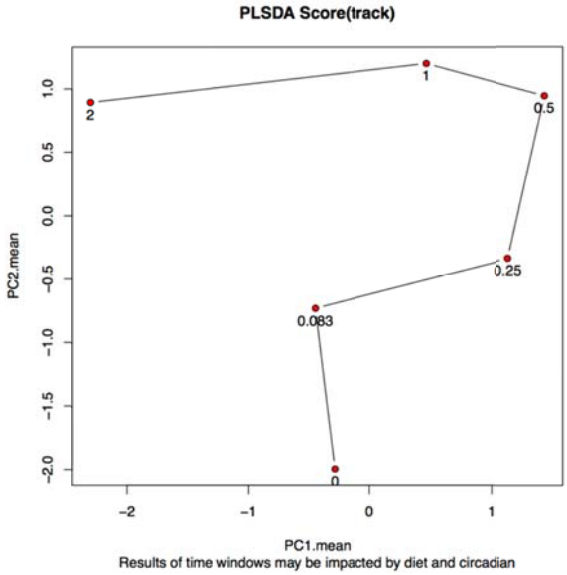

Figure S12. The track plot of PLS-DA.

HeatMap

A function to plot the heat map and clusters of input data

Arguments

data	The data under analysis (data.frame with required format). Please see **Fig.S1**, **Fig.S2** for detailed format.
cluster	A string indicating whether or in which direction the dendrograms should be drawn (“none”, “row”, “column” or “both”). Default: "both'.
scale	A character indicating whether the data should be centered and scaled before analysis and in which (“none”, “row” or “column”) direction. Default: "row".
filepath	A character string indicating the path where the results may be saved in.
design	(optional) a study design dataset (data frame with required format). See **Fig.S4** for detailed format. Default: "FALSE"

Usage and results

```
data("A")
HeatMap(data=A,cluster="both",scale="row",filepath=getwd(),design)
```

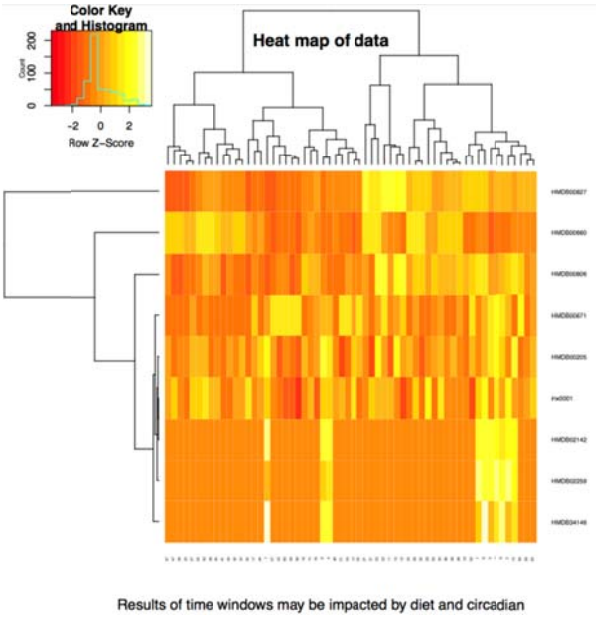

Figure S13. The heatmap plot.

Reference

Benjamini, Y.*, et al.* Controlling the false discovery rate in behavior genetics research. *Behav Brain Res* 2001;125(1-2):279-284.

Holm, S. A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics* 1979;6(2):65-70.

Jia, W.*, et al.* The polypharmacokinetics of herbal medicines. *Science* 2015;350(6262):S76-S79.

Khodakarim, S., Tabatabaei, S.M. and AlaviMajd, H. The multivariate nonparametric methods for identifying gene sets with differential expression. *Gene* 2014;552(1):18-23.

Lan, K. and Jia, W. An integrated metabolomics and pharmacokinetics strategy for multi-component drugs evaluation. *Curr Drug Metab* 2010;11(1):105-114.

Lan, K., Xie, G. and Jia, W. Towards polypharmacokinetics: pharmacokinetics of multicomponent drugs and herbal medicines using a metabolomics approach. *Evid Based Complement Alternat Med* 2013;2013:819147.

Lazar, C.*, et al.* Accounting for the Multiple Natures of Missing Values in Label-Free Quantitative Proteomics Data Sets to Compare Imputation Strategies. *J Proteome Res* 2016;15(4):1116-1125.

Troyanskaya, O.*, et al.* Missing value estimation methods for DNA microarrays. *Bioinformatics* 2001;17(6):520-525.

Tusher, V.G., Tibshirani, R. and Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* 2001;98(9):5116-5121.

Wei, R.*, et al.* GSimp: A Gibbs sampler based left-censored missing value imputation approach for metabolomics studies. 2017.

Xie, G.*, et al.* Poly-Pharmacokinetic Study of a Multicomponent Herbal Medicine in Healthy Chinese Volunteers. *Clin Pharmacol Ther* 2017.

Xie, G.*, et al.* Metabolic fate of tea polyphenols in humans. *J Proteome Res* 2012;11(6):3449-3457.