Introduction of example files

**R packages used in this paper:**

**R package “vegan” for variation partitioning analysis**

***Code:*** variation\_partitioning\_analysis.R

***Example file:***

M\_f3.csv,

M\_n.csv

***Introduction:***

M\_f3.csv→Epidemiological factors of each individual.

M\_n.csv→The log 10 transformed concentrations of high-frequency chemicals, here only 4 chemicals were showed.

**R package “FactoMineR” for Principal component analysis**

***Code: PCA.R***

***Example file:*** PCA.xlsx

***Introduction:*** Group and concentrations variables were used. Group depends on the epidemiological factors to be analyzed. Concentrations of high-frequency chemicals were log 10 transformed, here only 4 chemicals were showed. column “group” represents color and circle.

**R package "MatchIt" for Propensity Score Matching**

***Code:*** PSM.R

***Example file:*** hyperlipidemiaControl.xlsx

***Introduction:*** A file including epidemiological factors and disease conditions that you want them to be balance in disease and control group. Here, taking hyperlipidemia for example, the excel file includes column “hyperlipidemia” and other 13 confounders need to match.

**R package "RSC" for dose-risk relationship analysis**

***Code:*** RCS.R

***Example file:*** hyperuricemia\_RCS.xlsx

***Introduction:*** A data frame contains 3 kinds of variables. First column “hyperuricemia” represents the condition of hyperuricemia or normal; second epidemiological factors are the confounders need to be adjusted; finally, the chemicals represent the target of the desired dose-risk relationship, the concentrations were log 10 transformed (taking 2 chemicals for example).

**R package " WQS " for weighted quantile sum regression of mixture effect analysis**

***Code:*** WQS.R

***Example file:*** hyperuricemia\_mixture\_effect.xlsx

***Introduction:*** A data frame contains 3 kinds of variables. First column “group” represents the condition of hyperuricemia or normal; second, epidemiological factors are the confounders need to be adjusted; finally, the chemicals represent the target of the desired mixtures effect, the concentrations were log 10 transformed and scaled (taking 4 chemicals for example).

**R package " q g-comp " for quantile g-computation of mixture effect analysis**

***Code:***qg-comp.R

***Example file:*** hyperuricemia\_mixture\_effect.xlsx

***Introduction:*** Same as above.

**R package " BKMR " for Bayesian Kernel Machine Regression of mixture effect analysis**

***Code:*** BKMR.R

***Example file:*** hyperuricemia\_mixture\_effect.xlsx

***Introduction:*** Same as above

**Codes used for plot in this paper:**

**sunburst plot**

***Code:*** plotly\_sunburst.Rmd

***Example file:*** sunburst\_267chemicals.csv

***Introduction:*** There are 3 columns in the csv file. Column “labels” represents the label of chemicals of the outermost circle. Column “parents” represents which group the “labels” belongs to. Column “values” represents the number of chemicals in each group.

**bar plot**

***Code:*** bar\_plot.R

***Example file:*** Variation\_Proportion.csv

***Introduction:*** The first column is the names of each group, the second group is the value of each group.

**stacked bar plot (flip)**

***Code:*** stacked\_bar\_plot(flip).R

***Example file:*** location\_74\_15provience.csv

***Introduction:*** A matrix of geometric mean concentration of chemicals in human from different provinces. Rows represent provinces and columns represent classes of chemicals.

**line & point & error bar plot**

***Code:*** point&line&errorbar\_plot.R

***Example file:*** age\_increase.xlsx

***Introduction:*** There are 3 columns in the excel file. Column “Concentration” represent the scaled concentration of chemicals in human serum. “Exposures” column represents the chemical`s name. Column “Age” represents different age ranges.

**split violin plot**

***Code:*** split\_violin\_plot.R

***Example file:*** gender\_beta.HCH.xlsx

***Introduction:*** There are 3 columns in the excel file. Column “Concentration” represent the concentration of chemicals in human serum. Column “Exposures” represents the chemical`s name. Column “Gender” represents different gender groups.

**violin & box plot**

***Code:*** Violin&Box\_plot.R

***Example file:*** clinic\_info\_uricAcid.csv

***Introduction:*** The first column represent the class, and the second column represent the values of each individual in the class.

**forest plot**

***Code:*** forestplot\_range.R

***Example file:*** hyperlipidemia\_OR\_log10.csv

***Introduction:*** The column “exposures” represents the name of each chemical. the column “OR” represents the odd ration “LowCI” represents the lower limit of confidence interval, “HighCI” represents the higher limit of confidence interval, “Pvalue” represents the significant of two tails. “OR”, “LowCI”, “HighCI”and“Pvalue”are all derived from the binary logistic regression model, the data is log 10 transformed concentrations.

**stacked bar plot**

***Code:*** stack\_bar\_plot.R

***Example file:*** 9clinicinfo.csv

***Introduction:*** A matrix of -1, 0 and 1. Rows represent clinic informations and columns represent chemicals. -1, 0 and 1 represent the negative, none and positive association between clinic information and chemicals.

**heatmap**

***Code:*** heatmap&significant.R

***Example file:***

hyperlipidemia\_age\_OR.csv

hyperlipidemia\_age\_p.csv

***Introduction:***

hyperlipidemia\_age\_OR.csv→ A matrix of odd ratios derived from the binary logistic regression model. Rows represent chemicals and columns represent 3 age groups.

hyperlipidemia\_age\_p.csv→A matrix of *p* values derived from the binary logistic regression model. Rows represent chemicals and columns represent 3 age groups.

**jitter plot**

***Code:*** jitter\_plot.R

***Example file:*** BE&HMB2&BGV.csv

***Introduction:*** The first column represents the class of chemical. The second column represents hazard quotients of the class for each individual.