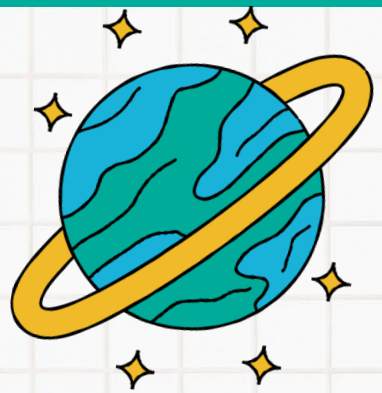
【生物資訊系列】6

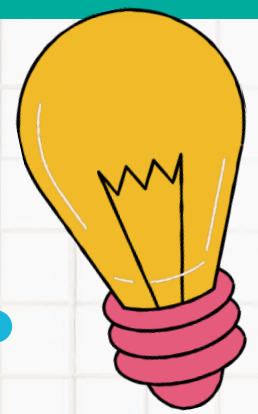
Metabolomics data analysis

Yu-Jen Liang
Dr. Hsin-Chou Yang's Lab, ISS

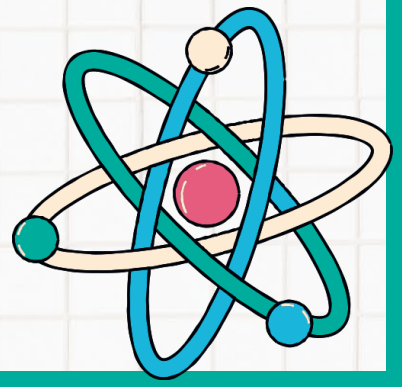
2024.5.14



INTRODUCTION



Metabolomics data analysis is a field within the broader realm of omics sciences, which focuses on the comprehensive study of small molecules, or metabolites, within a biological system. These metabolites include a wide range of compounds such as sugars, lipids, amino acids, and organic acids, among others. Metabolomics aims to provide a snapshot of the metabolic state of an organism, tissue, or cell at a given point in time, reflecting the dynamic interplay of various biochemical pathways and environmental factors.



01

BASIC CONCEPTS

News events

食安

多氯聯苯:

米糠油中毒事件

3級致癌物蘇丹紅:

辣椒粉

邦克列酸:

寶林茶室食物中毒

藥物濫用

毒品或迷幻物品:

安非他命海洛因、大麻、古柯鹼、安眠鎮靜劑

餵藥事件:

苯巴比妥、苯二氮平類

環境污染

溴化阻燃劑:

降低或抑制材料易燃性，建築材料、電子與電器產品及紡織品。

全氟烷基磺酸與羧酸類化合物:

疏水與疏油的特性，普遍應用於地毯防污劑、防火劑、冷卻劑與殺蟲劑。

過氯酸根離子:

農業用化學肥料中亦含有過氯酸根成分，污染地表水與地下水。

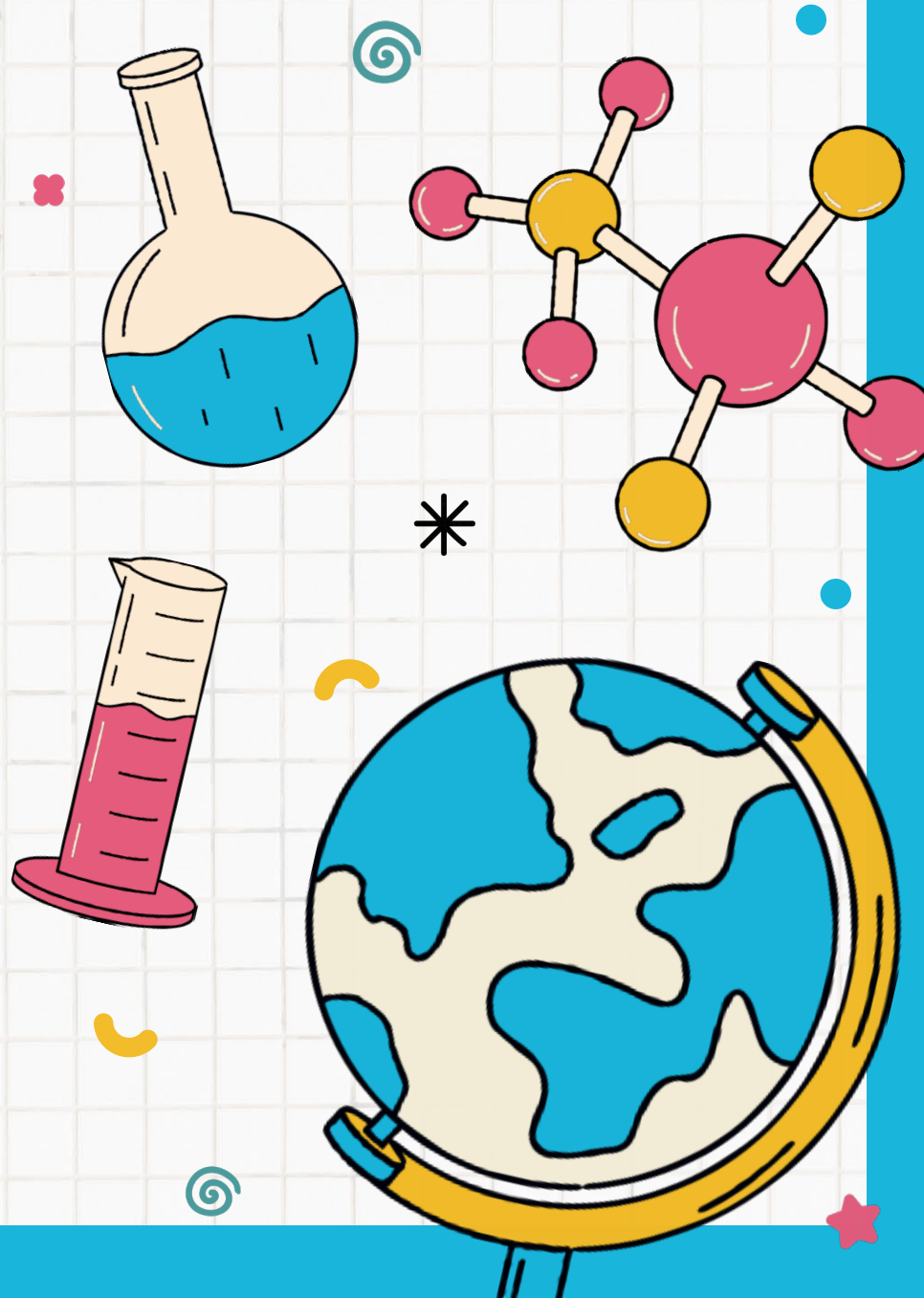
食品用藥

動物、植物、加工品:

抗生素、防腐劑、農藥、放射線物質(碘、銫)、萊克多巴胺、瘦肉精、乙型受體素

02

OBJECTIVES



GOALS

代謝體研究與食品安全、藥物使用、環境污染和毒品等之間存在著多重關聯性：

食品安全：

代謝體研究可以幫助評估食品中的潛在有害物質，如農藥殘留、重金屬等。通過分析人體代謝產物的變化，可以識別食品暴露的個體，評估其對健康的影響。

藥物使用：

代謝體研究可用於瞭解藥物在人體內的代謝途徑及代謝產物，進而理解藥物的作用機制、個體對藥物的反應以及藥物的副作用。這有助於個體化治療和藥物安全性的評估。

環境污染：

代謝體研究可以作為評估環境污染對生物體的影響的一種方法。通過分析生物體內的代謝物變化，可以識別暴露於環境污染物中的個體，並評估其對健康的潛在風險。

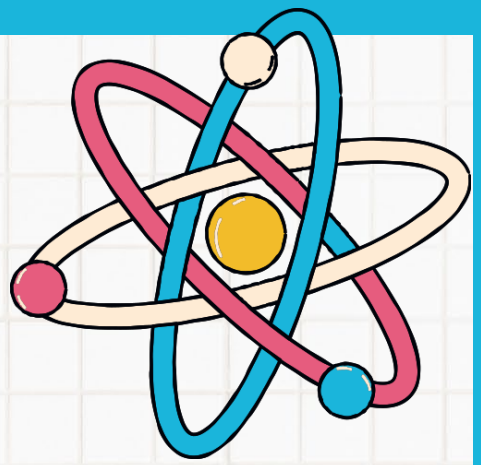
毒品濫用：

代謝體研究也可應用於研究毒品對人體的影響。通過分析毒品使用者的代謝物，可以了解毒品的代謝途徑、對器官組織的影響以及潛在的健康風險。

代謝體研究在多個領域中都發揮著重要作用，有助於深入理解生物體對內外環境的反應，從而促進食品安全、藥物安全和環境保護等方面的工作。

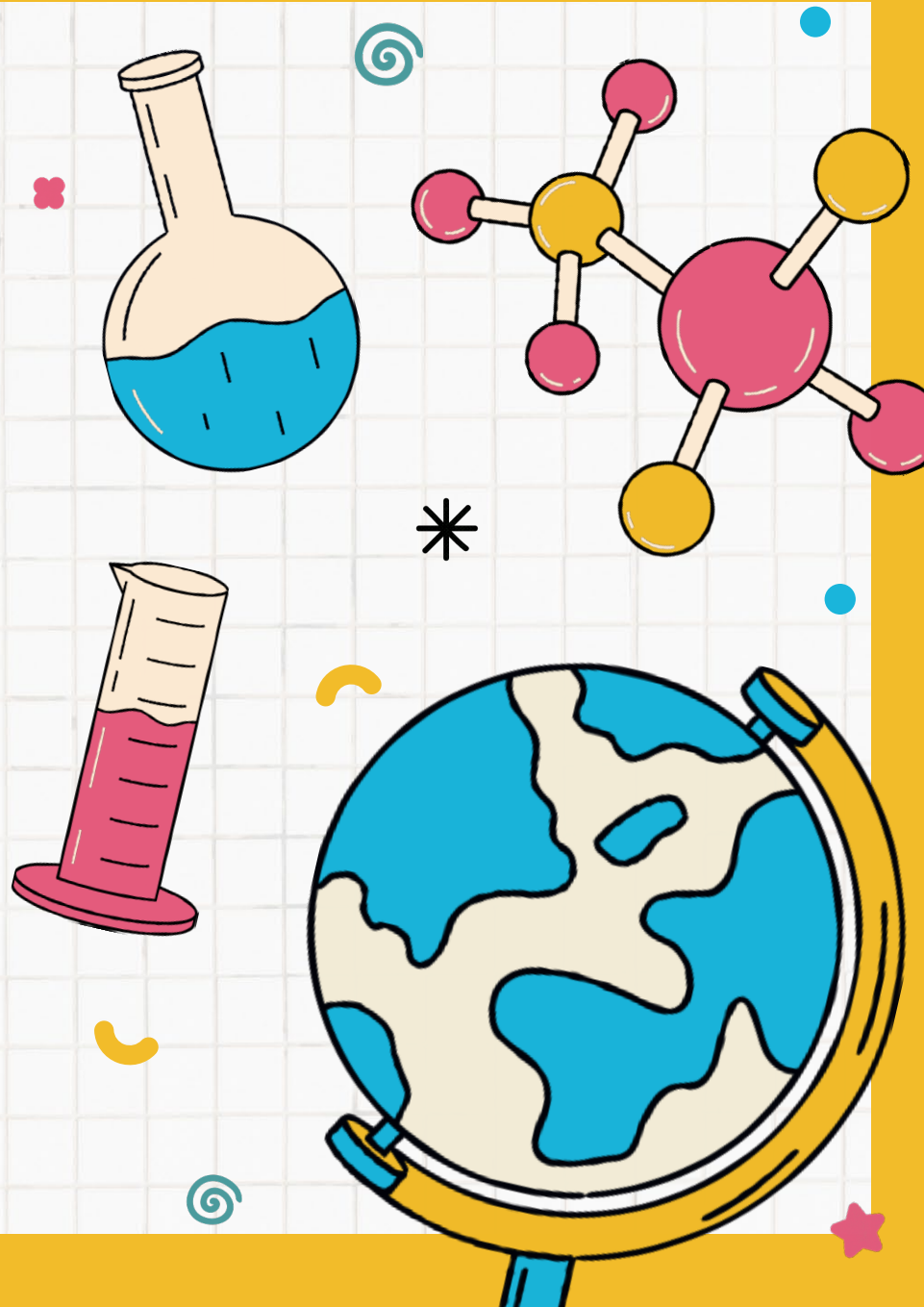
Outline

- Introduction
- Background and motivation
- Our statistical metabolomics software -- SMART
- Data demonstration
  - Antihypertensive pharmacometabolomics study
  - Drug data
  - Breast cancer data
- Conclusion and discussion

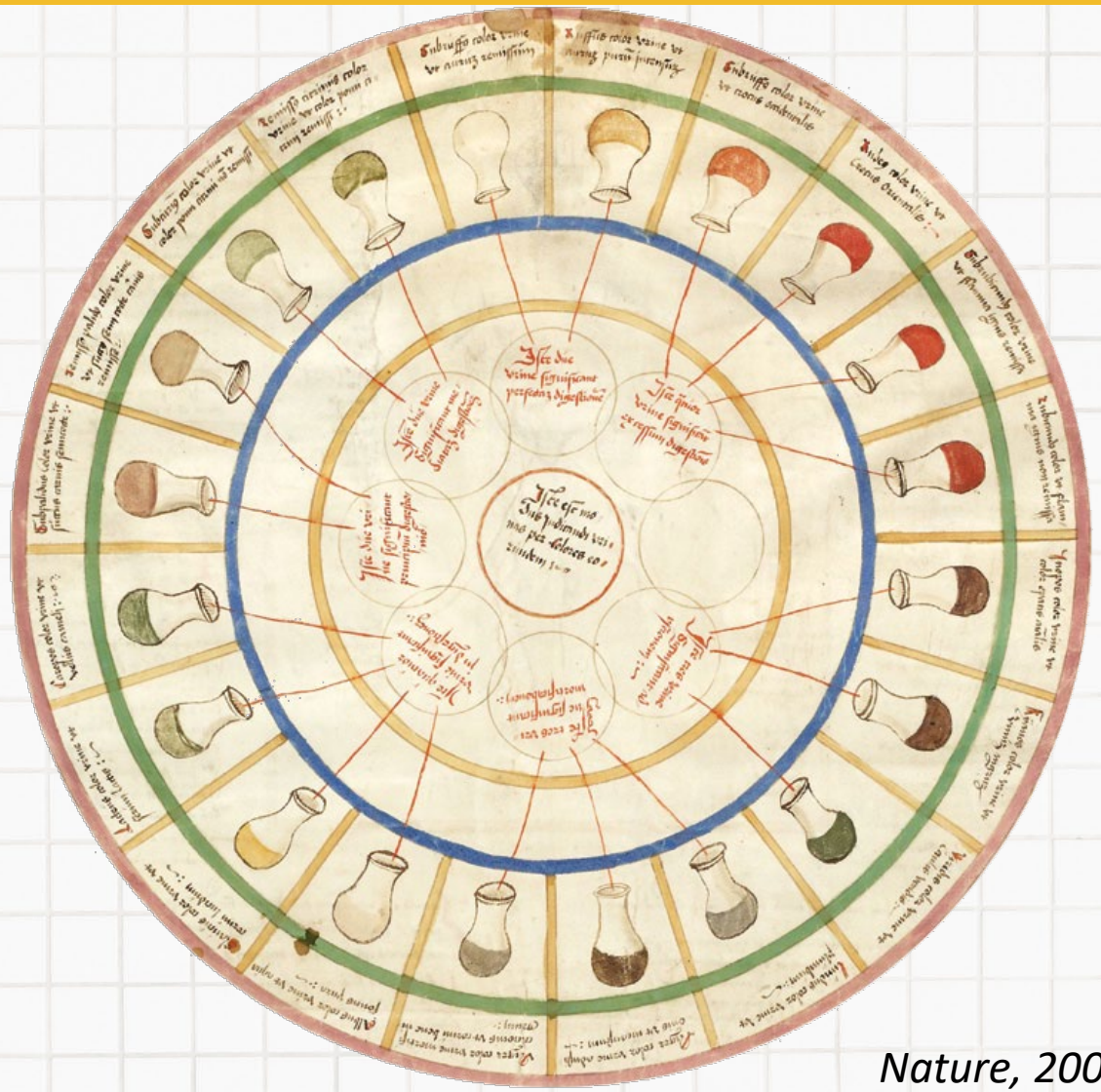


03

Introduction



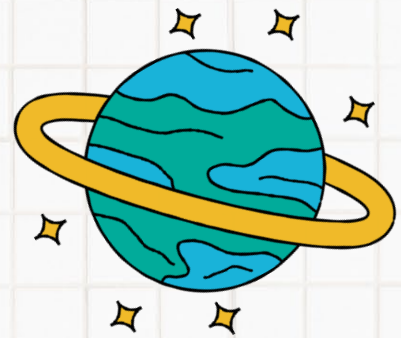
1506年 古希臘學者 Pinder 所發表的 尿輪

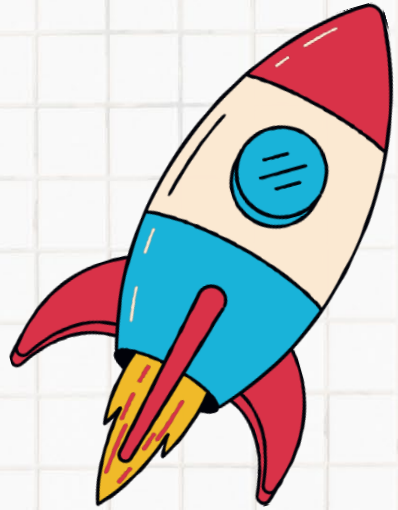


Nature, 2008

This **urine wheel** was published in 1506 by Ullrich Pinder, in his book *Epiphanie Medicorum*. It describes the possible colours, smells and tastes of urine, and uses them to diagnose disease.

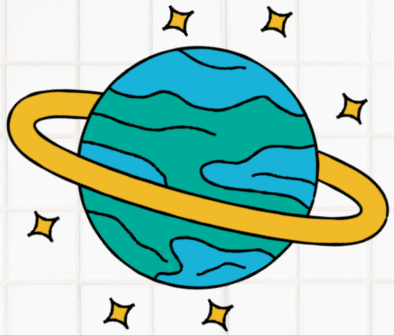
- Metabolite
- Metabolome
- Metabolomics





Metabolite

Metabolites are small molecules involved in metabolism processes [Wishart, 2007].

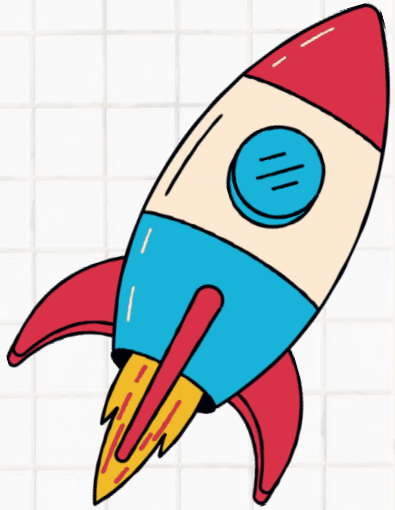


Endogenous metabolites

Amino acids, organic acids, nucleic acids, fatty acids, amines, sugars, vitamins, co-factors, pigments, and antibiotics are **produced** by organisms in the body **naturally**.

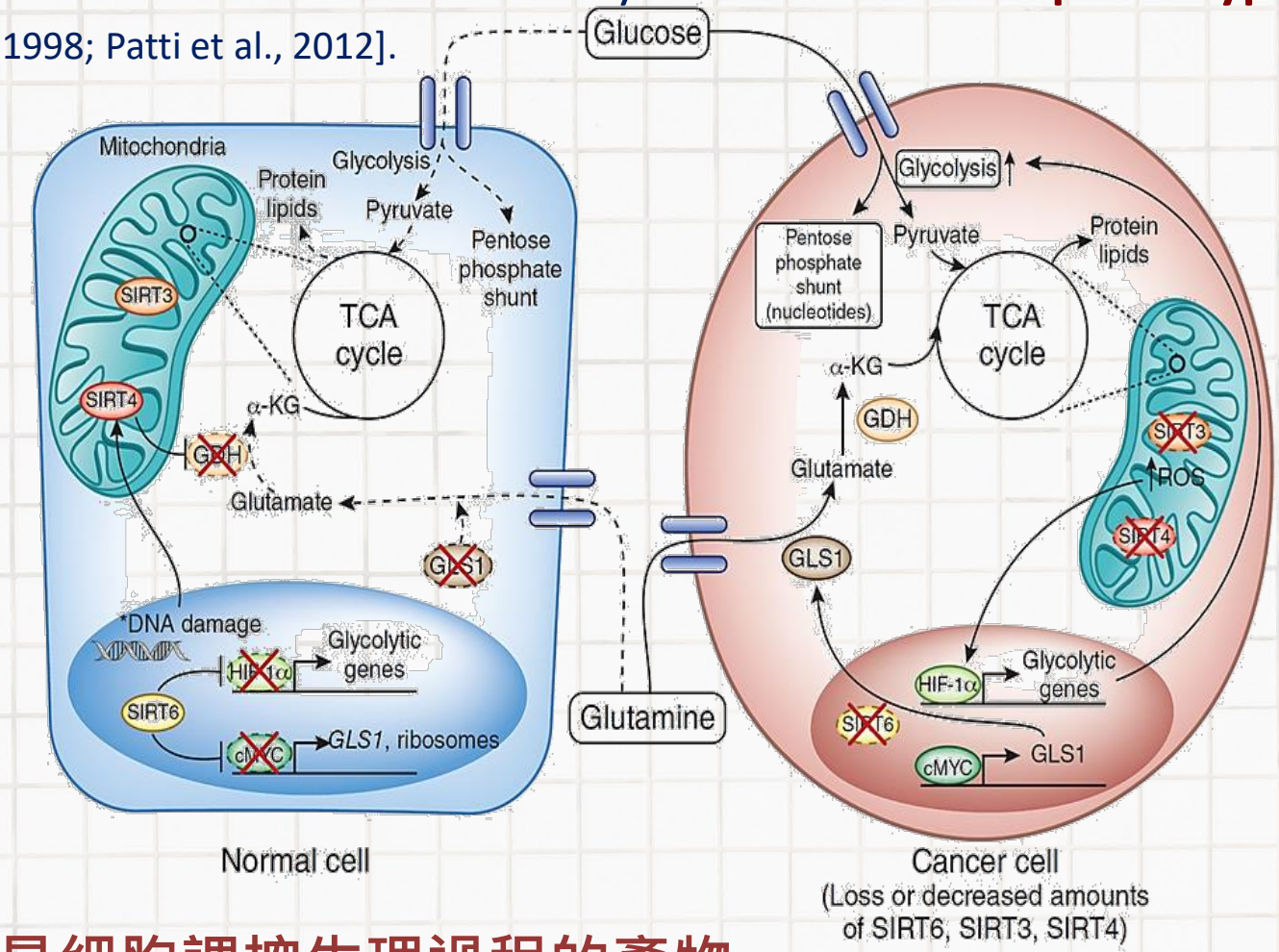
Exogenous chemicals

Drugs, environmental contaminants, food additives, toxins, and other xenobiotics are derived **externally**.



Metabolome (1)

Metabolome, typically defined as a collection of **all metabolites** produced by cells, offers a window to investigate cellular biochemistry and how the mechanistic biochemistry relates to **cellular phenotype** [Oliver et al., 1998; Patti et al., 2012].

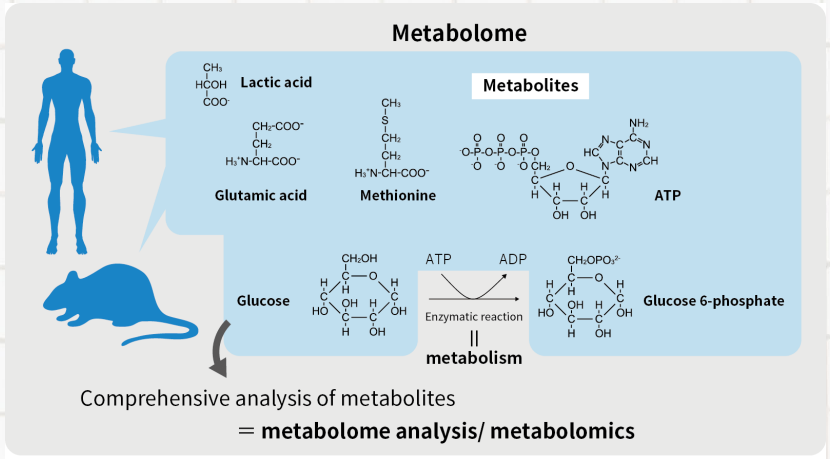


代謝物是細胞調控生理過程的產物

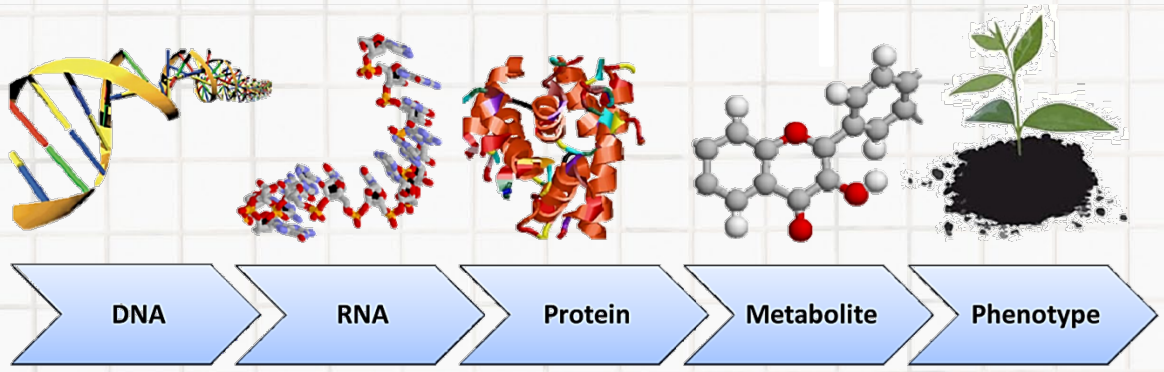
Nature Medicine, 2014

Metabolome (2)

- In comparison with **genome**, **transcriptome**, and **proteome**, **metabolome** is closer to downstream phenotypes and therefore serves as promising materials to study phenotypes.



https://humanmetabolome.com/ap/techinfo/



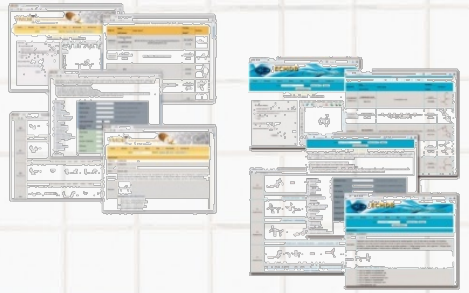
Genome & epigenome | Transcriptome & Exome | Proteome & Interactome | Metabolome & Fluxome | Phenome & Exposome

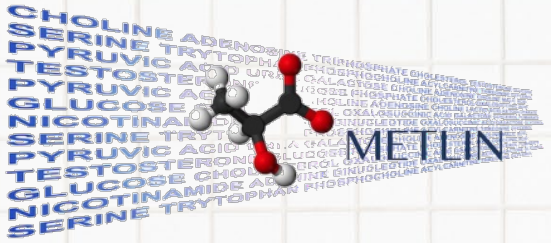
Int. J. Mol. Sci. **2013**

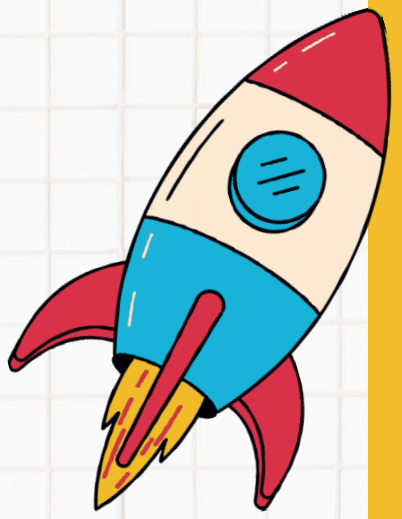
Metabolome (3)

- Because metabolome is also largely determined by genome, different **species** have different metabolomes.

- Metabolome databases of some species have been constructed
  - Yeast Metabolome Database (YMDB)
  - E. coli Metabolome Database (ECMDB)
  - Human Metabolome Database (HMDB)
  - MassBank Database
  - METLIN (METabolite LINk) Metabolomics Database

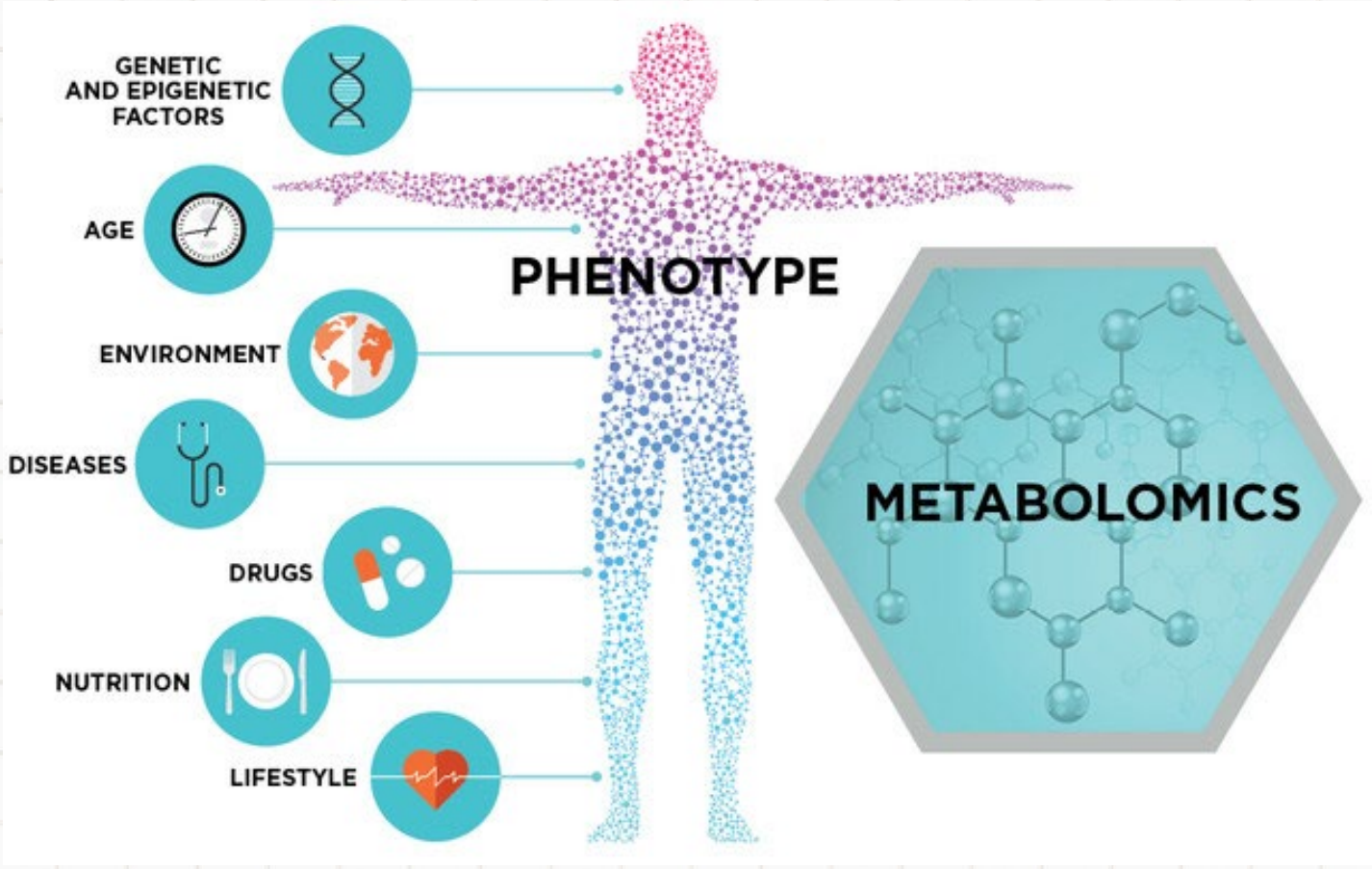






Metabolomics (1)

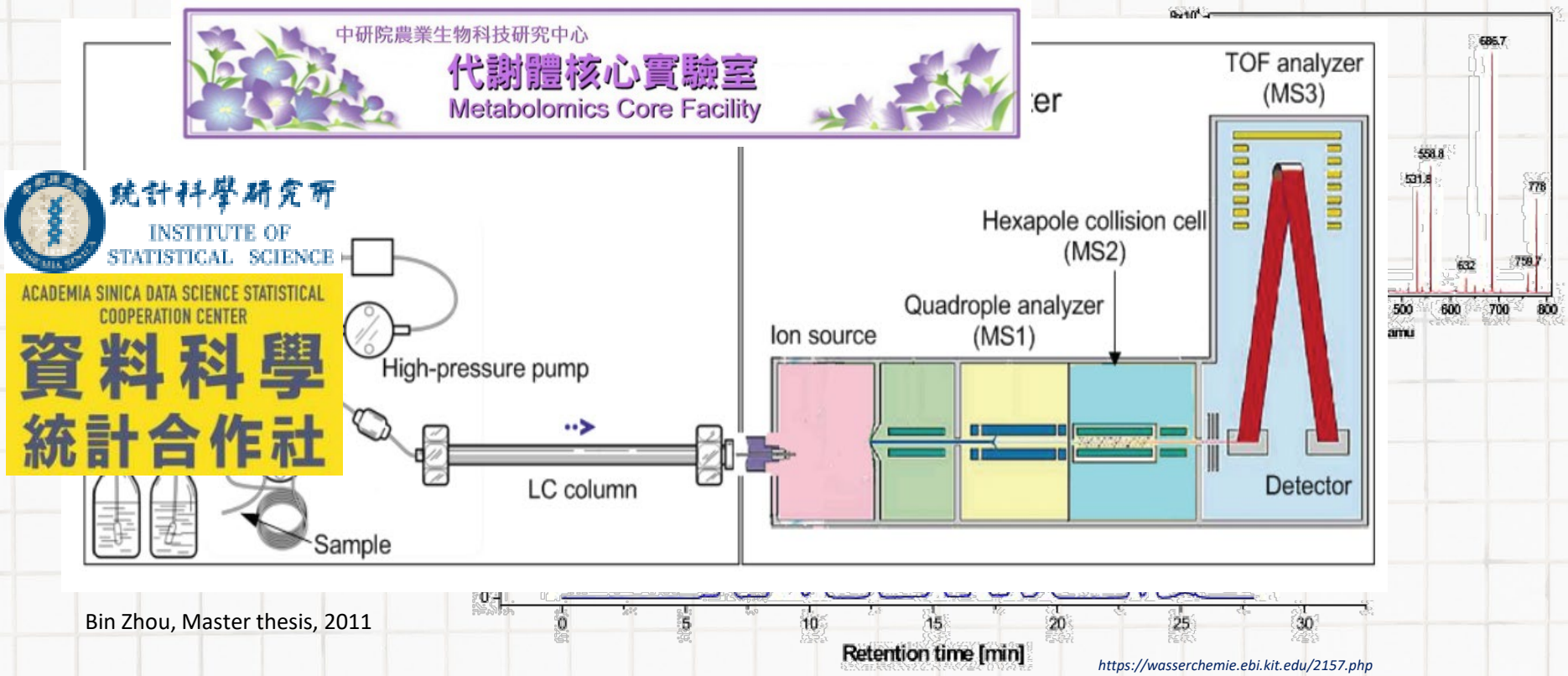
- Metabolomics was developed in the early 2000s as one of the OMICS tools.



https://www.mtidx.com/our-technology/metabolomics

Metabolomics (2)

- LC-MS and GC-MS are the most popular experiment platforms for metabolomics studies.
  - High sensitivity and selectivity in rapidly separating and quantifying target or non-target compounds, including metabolites and lipids, in biological fluids [Hird et al., 2014].



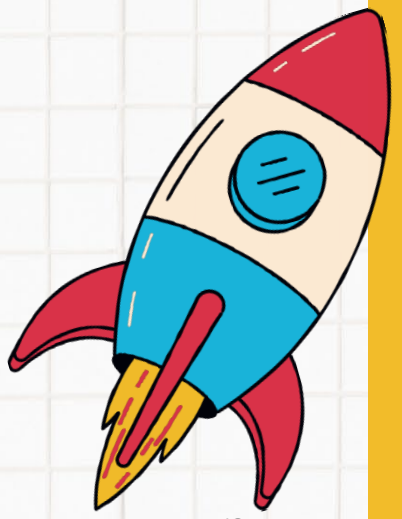
Bin Zhou, Master thesis, 2011

https://wasserchemie.ebi.kit.edu/2157.php

Metabolomics (3)

- Recently, **metabolomics profiling** has been used broadly in exploring
  - The association with **disease states** [Roede, 2013; Buergel, 2022]
  - **Drug metabolism** [Liu, 2007; Balashova, 2018]
  - **Clinical diagnostics** [Patti, 2012; Thistlethwaite, 2022]
  - Understanding mechanisms of **toxicity** [Clarke, 2008; Olesti, 2021]

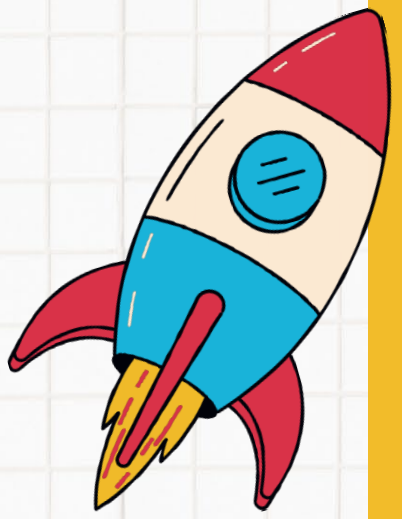
- The results have aided to understand the roles of metabolites in the biological mechanisms [Altuntas, 2014; Donia, 2015; Senyilmaz, 2015; Qiu, 2023]



Metabolomics (4)

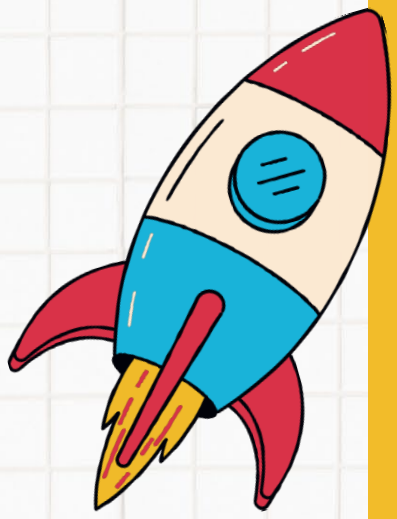
- As the popularity of metabolomics studies, a number of analysis tools are available.

- First, some **commercial software** for metabolomics data processing, peak alignment, and/or data visualization are provided with the experimental instruments such as
  - **Mass Profiler Pro** (Agilent Technologies, Palo Alto, CA, USA)
  - **ProfileAnalysis** (Bruker, Billerica, Massachusetts, USA)
  - **MarkerLynx** (Waters, Milfold, MA, USA)
  - **Compound Discoverer** (Thermo Fisher, Waltham, Massachusetts , USA)



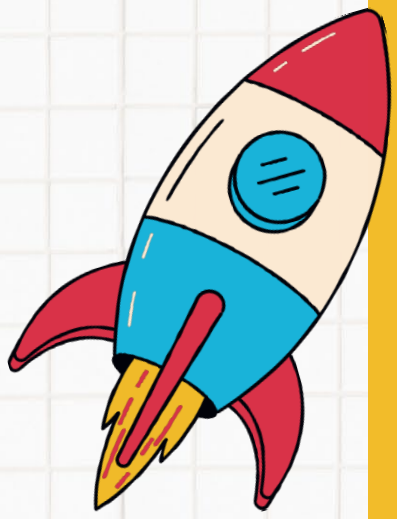
Metabolomics (5)

- Second, **free packages** were also developed
  - **MZmine** [Katajamaa, 2006]
    - A modular framework for MS-based data processing, visualization, and analysis.
  - **XCMS** [Smith, 2006]
    - MS-based data processing using a nonlinear peak detection, matching, and RT alignment.
  - **MultiAlign** [LaMarche, 2013]
    - An alignment of multiple LC/MS datasets by clustering mass and chromatographic elution features across datasets.



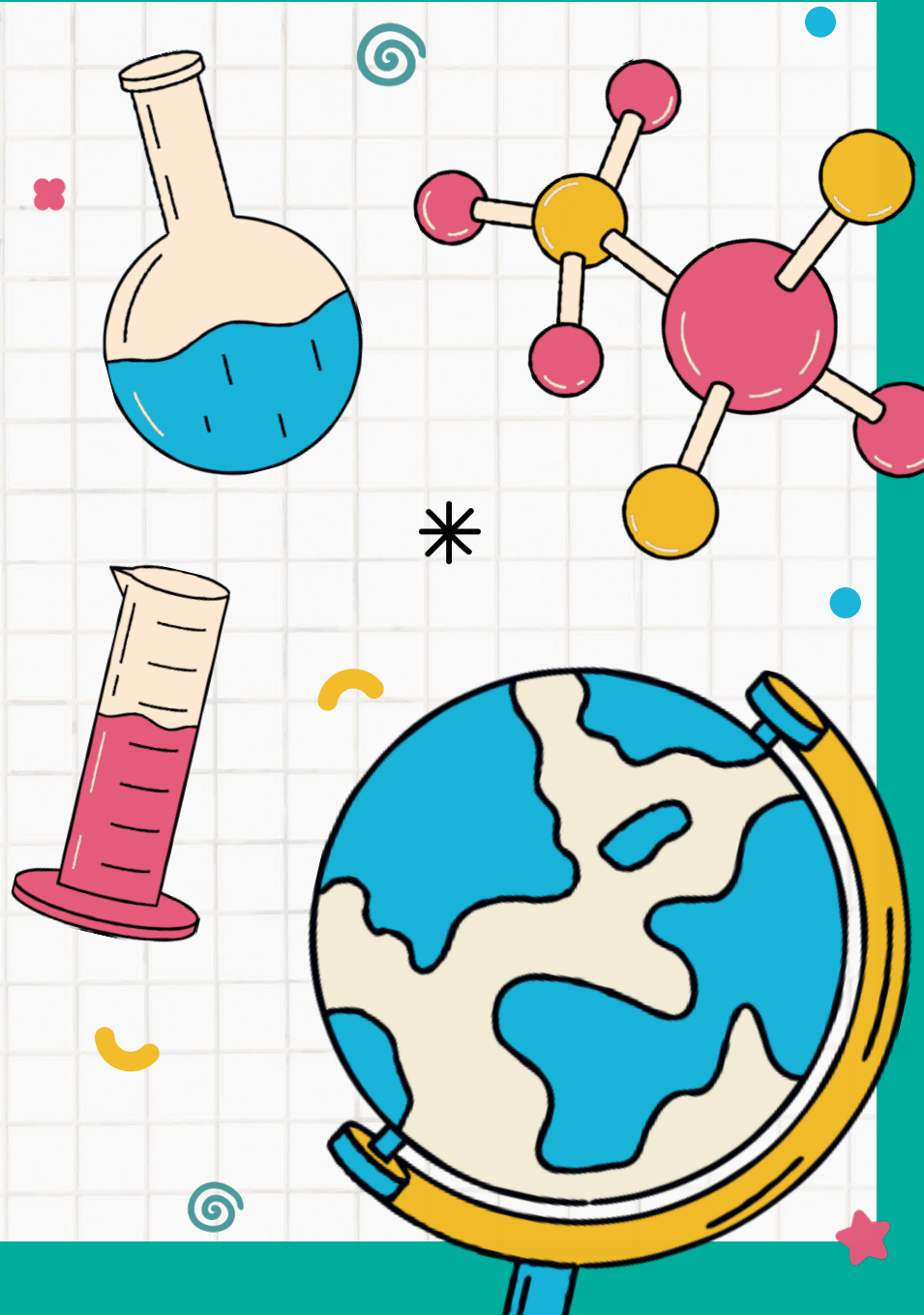
Metabolomics (6)

- **Web-based tools** are available for performing statistical analysis and pathway analysis for metabolomics data
  - **MetaboAnalyst** [Xia, 2015]
    - A web server designed to certificate comprehensive metabolomics data analysis, visualization, and interpretation.
  - **MeltDB** [Kessler, 2013]
    - A web application about data storage, sharing, standardization, integration, and analysis of metabolomics experiments.
  - **XCMS Online** [Tautenhahn, 2012]
    - A platform for feature detection, retention time (RT) correction, alignment, annotation, statistical analysis, and data visualization.



04

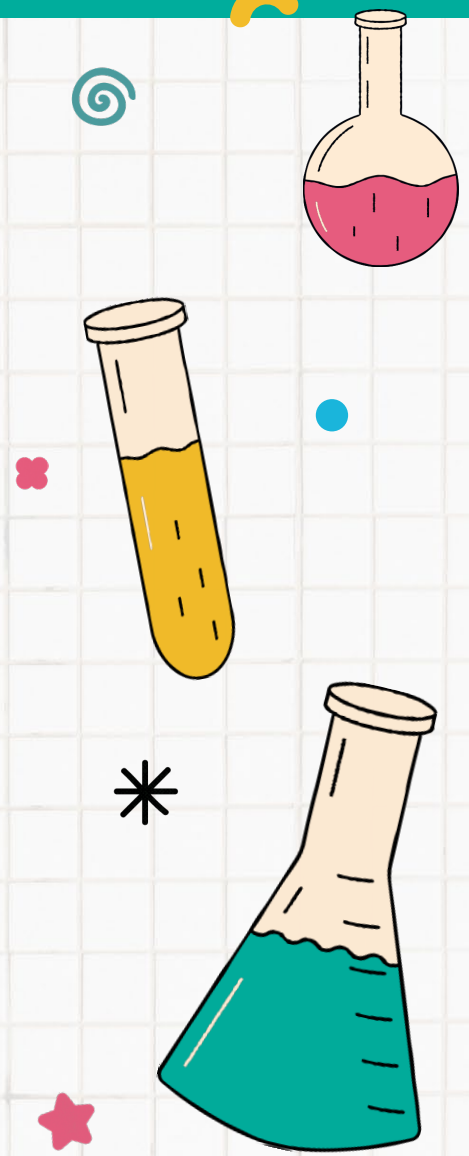
Background and motivation



Background (1)

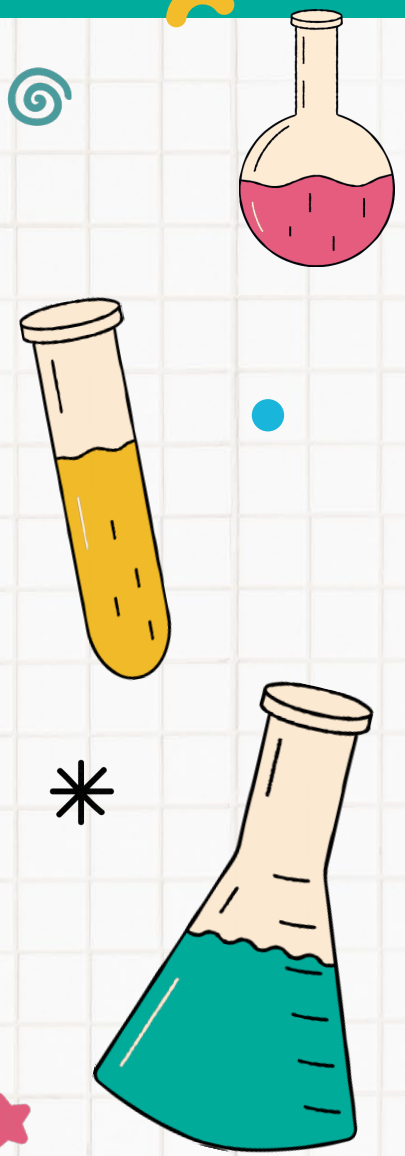
- Recently, numerous large-scale metabolome-wide association studies (**MWAS**) have been conducted [Altmaier, 2014; Osborn, 2013; Sekula, 2015; Tzoulaki, 2014; Zamora-Ros, 2014; Kojouri, 2023].

- Such studies are characterized by a **long study time**, complex **experimental factors**, and a **high number** of samples with multiple technical and biological replicates [Fernie, 2011].



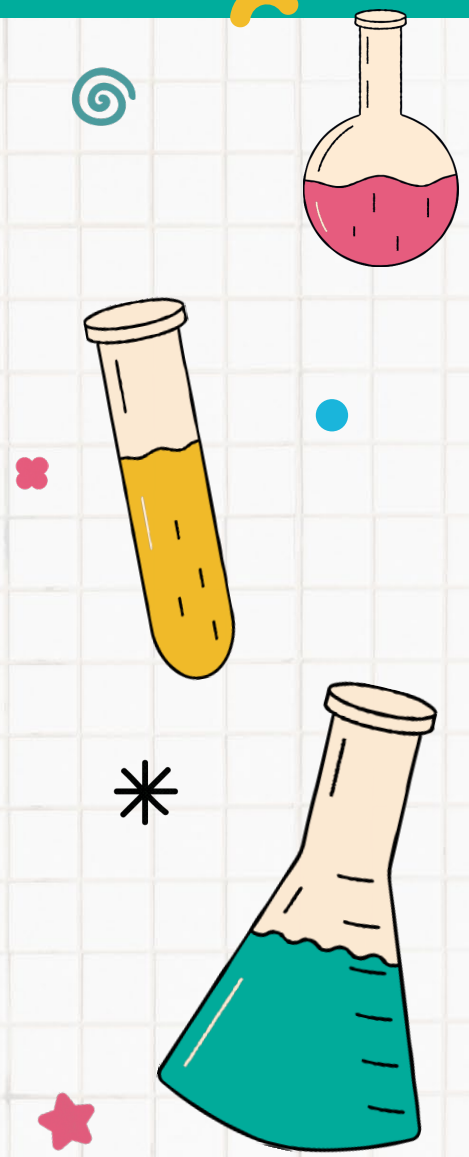
Background (2)

- The experiments of large-scale MWAS may be influenced by several sources of variation, including controllable and uncontrollable factors.
  - **Controllable factors**
    - **Sample characteristics** and **known batch effects**, can be considered during study design.
  - **Uncontrollable factors**
    - **Specimen contamination** and **unknown batch effects** (e.g., a batch effect due to period-to-period variation in LC column performance [Karpievitch, 2014]), should be appropriately addressed during statistical analysis.



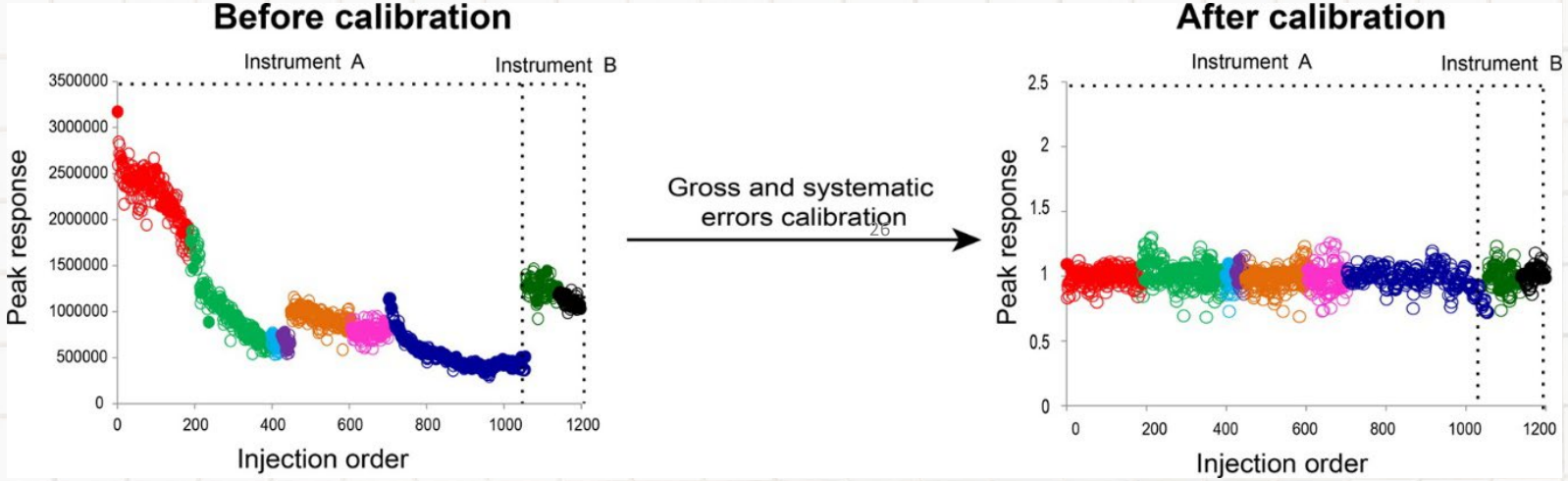
Background (3)

- Without appropriate data quality control and statistical modelling, false positives and negatives in statistical associations **increase** substantially.



Background (4)

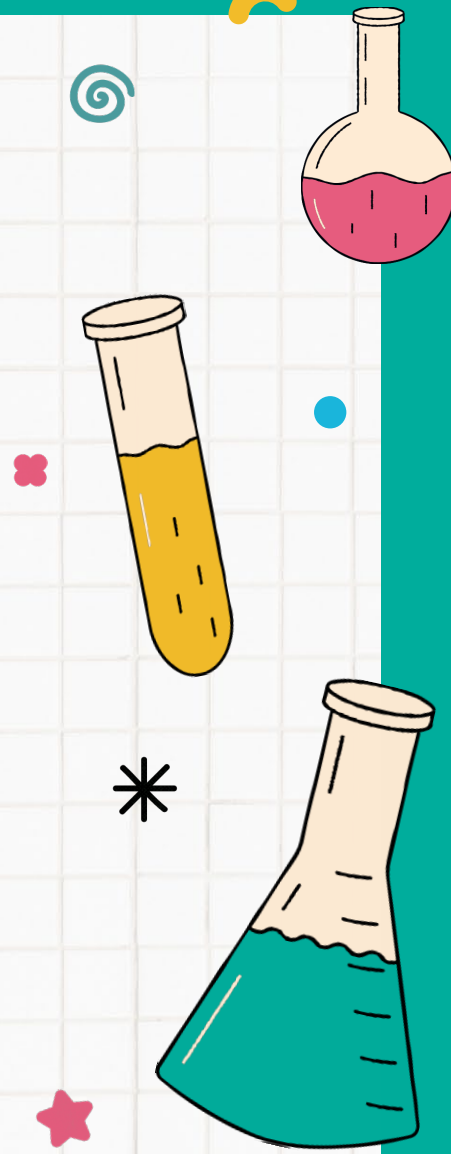
- **Data quality issues** and complex **batch effects** in a metabolomics study must be tackled by statistical analysis properly.



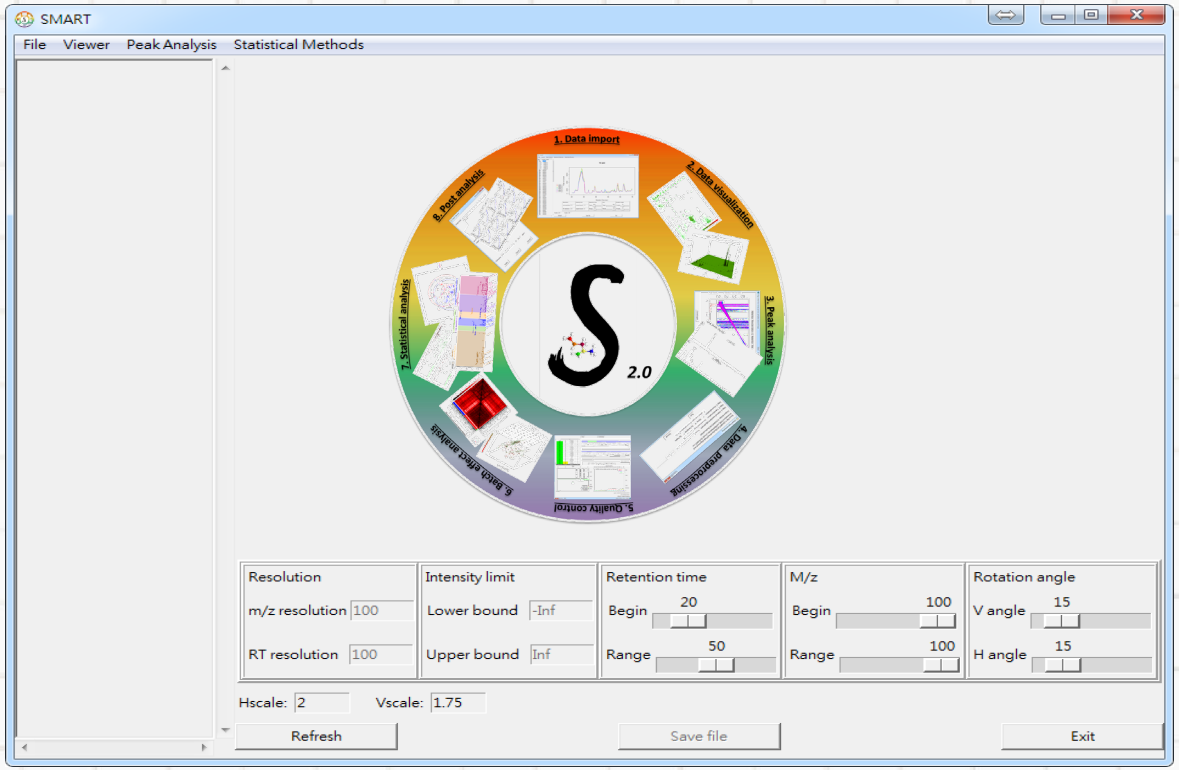
Anal. Chem. 2016

Motivation

- **No** software is available for a complete analysis:
  (1) analyzing different data file formats
  (2) visually representing various types of data features
  (3) implementing peak alignment
  (4) conducting quality control
  (5) exploring batch effects
  (6) performing association analysis
  (7) accomplishment peak identification
- This study is aimed to develop **an integrated analysis tool** for statistical metabolomics studies to streamline the **complete analysis flow** from initial data preprocessing to downstream association analysis.



SMART: Statistical Metabolomics Analysis – An R Tool



SMART with a user-friendly interface was developed in R and R-GUI (Graphical User Interface) under the Windows operating system.

代謝體資料分析軟體

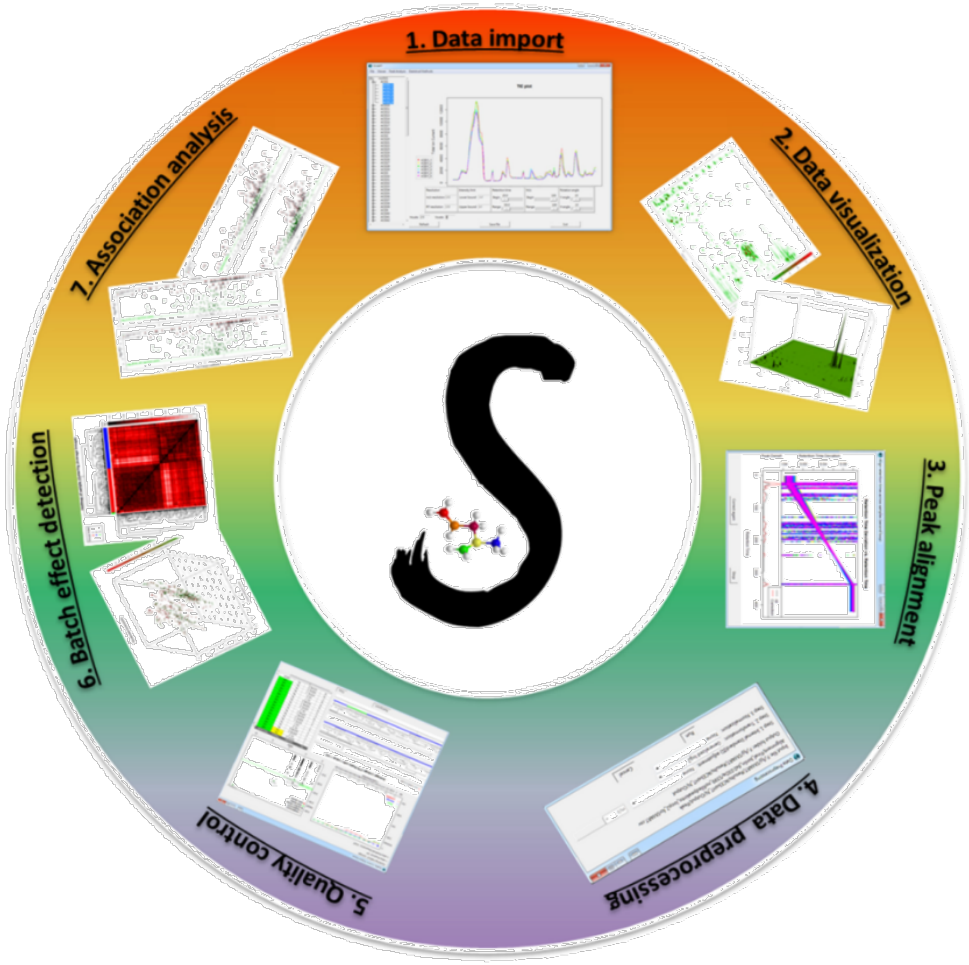
History & Work flow

SMART DESCRIPTION

Yu-Jen Liang and Chih-Ting Yang
Dr. Hsin-Chou Yang's Lab

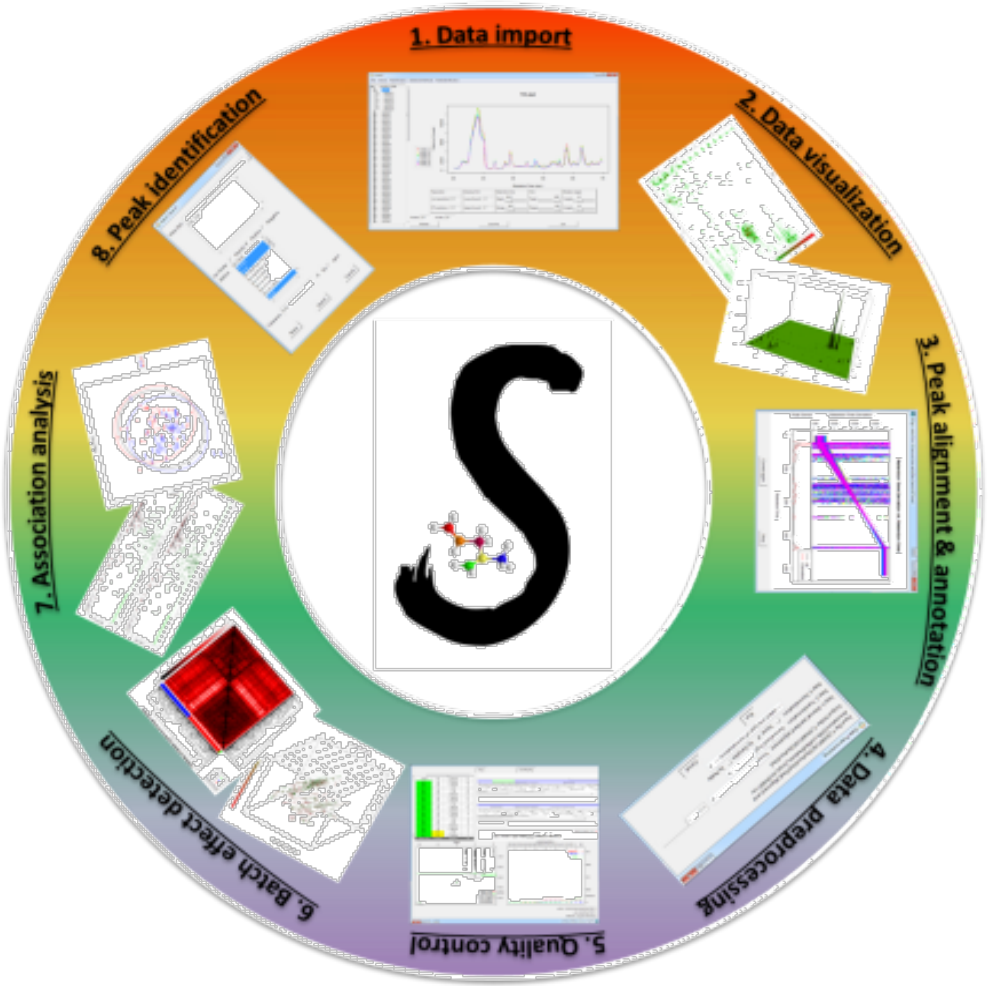
SMART description

- SMART (Statistical Metabolomics Analysis - an R Tool)
  - with a user-friendly interface was developed in R and R-GUI (Graphical User Interface) under the Windows and Mac operating system.



Version 1.2

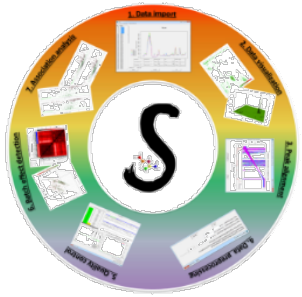
Statistical Metabolomics Analysis – an R Tool (SMART)



4. Data preprocessing
--- Pareto scaling (PS)
7. Association analysis
--- Partial least squares (PLS)
8. Peak identification
--- HMDB & Massbank

Website

http://www.stat.sinica.edu.tw/hsinchou/metabolomics/SMART.htm

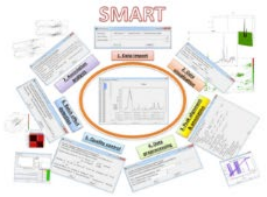


Version 1.0

SMART (Statistical Metabolomics Analysis - An R Tool)

Introduction:

SMART written in R and R GUI has been developed as user-friendly software for integrated analysis of metabolomics data. SMART streamlines the complete analysis flow from initial data preprocessing to downstream association analysis, including analyzing different data file formats (e.g., .raw, .d, and mzXML), visually representing various types of data features (e.g., total ion chromatogram and mass spectra), implementing peak alignment, conducting quality control for samples and peaks, exploring batch effects (e.g., known experimental conditions, unknown latent groups, or hidden substructures), and performing association analysis.



Software:

Version	System	Download File	File Type	File Size	Update Time
SMART 1.2 (beta)	Windows (64-bit)		7z	1.75 MB	2019/Mar/27
SMART 1.1	Windows (32-bit)		7z	75.0 KB	2016/Jul/26
	Windows (64-bit)		7z	79.7 KB	2016/Jul/26
	Mac		7z	48.6 KB	2016/Jul/26
SMART 1.0	Windows (32-bit)		7z	75.2 KB	2016/May/19
	Windows (64-bit)		7z	79.7 KB	2016/May/19
	Mac		7z	50.1 KB	2016/May/19

- Other required packages: (1) R language and packages, (2) ActiveTcl 8.5 (32-bit, 64-bit, Mac), and (3) ProteoWizard.
- Software for unzipping *.7z files can be downloaded from the website: http://www.7-zip.org/

Example: An antihypertensive pharmacometabolomics study of 5 nonmedication and 5 ACEi medication patients

Version	Description	Download File	File Type	File Size	Update Time
SMART	Raw spectrum files		7z	853 Mb	2016/May/19
	Covariate file		CSV	4 Kb	2016/May/19
	Factor file		CSV	4 Kb	2016/May/19

User guide:

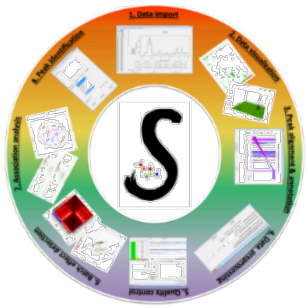
Version	Description	Download File	File Type	File Size	Update Time
SMART 1.2	Instructions		PDF	1.82 Mb	2019/Mar/27
SMART 1.1	Instructions		PDF	2.11 Mb	2016/Jul/26
SMART 1.0	Instructions		PDF	2.11 Mb	2016/May/19

Citation:

Yu-Jen Liang, Yu-Ting Lin, Chia-Wei Chen, Chien-Wei Lin, Kun-Mao Chao, Wen-Harn Pan and Hsin-Chou Yang (2016/05). SMART: Statistical Metabolomics Analysis – An R Tool. *Analytical Chemistry* **88**, 6334 – 6341.

Correspondence:

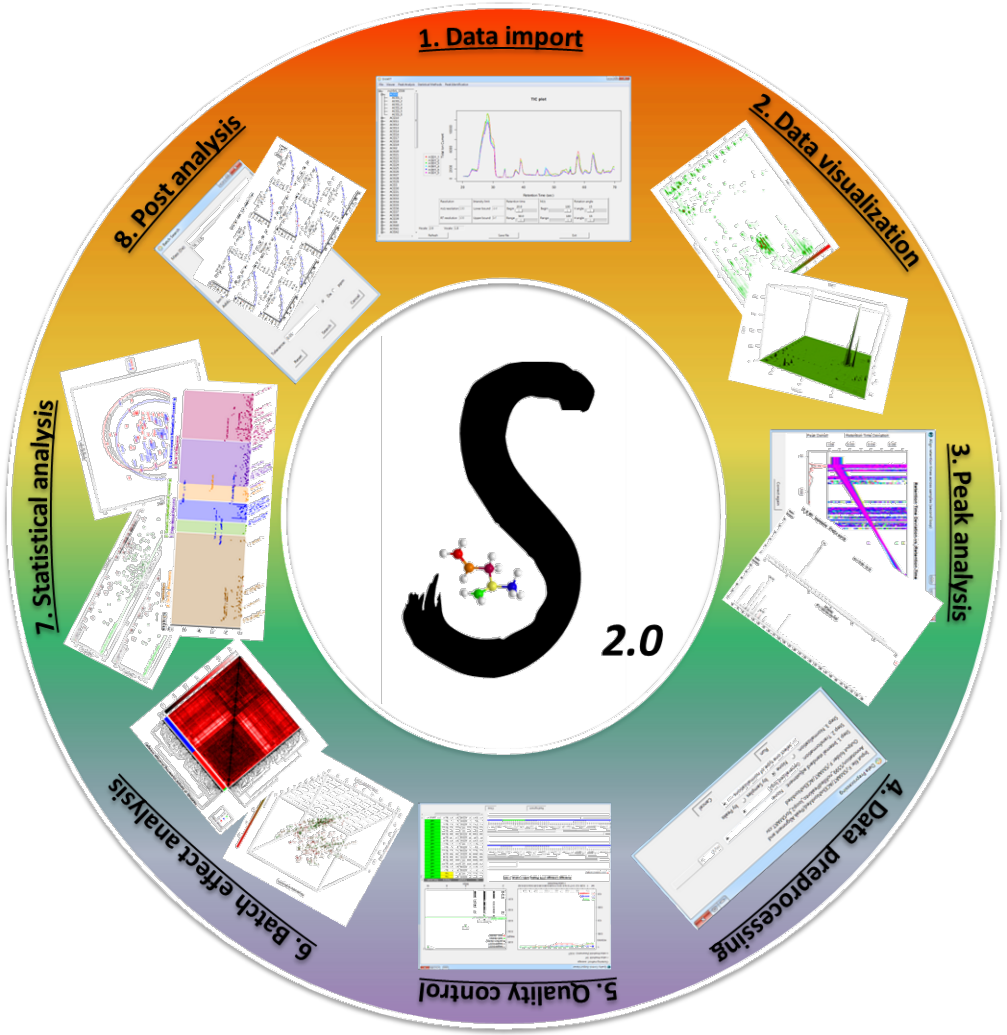
Dr. Hsin-Chou Yang, Institute of Statistical Science, Academia Sinica, 128, Academia Road, Section 2 Nankang, Taipei 115, Taiwan. (Fax) 886-2-27831523; (Tel) 886-2-27889311 ext. 113; (E-mail) hsinchou@stat.sinica.edu.tw



Version 1.2

Statistical Metabolomics Analysis – an R Tool (SMART)

Version 2.0



3. Peak analysis
- --- untargeted data analysis
- --- targeted data analysis

4. Data preprocessing
- --- Pareto scaling (PS)
- --- Inverse normal transformation

5. Quality control
- --- S/N ratio
- --- Clustering distance measures

7. Statistical analysis ~~(Association analysis)~~
- --- ANCOVA
- --- Partial least squares (PLS)
- --- Integrative omics pathway analysis (IOPA)

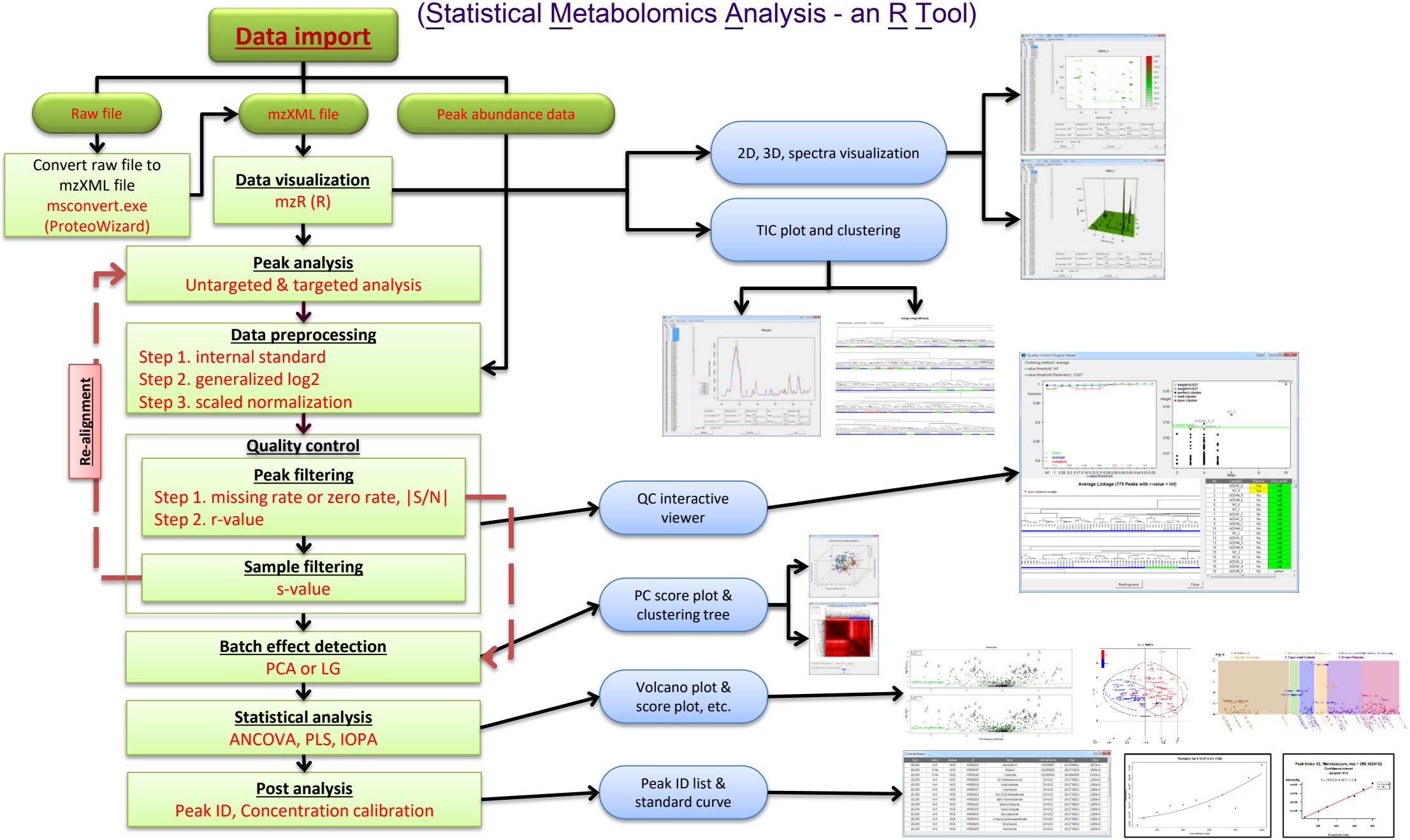
8. Post analysis

Peak identification
- --- HMDB & MassBank

Concentration calibration
- --- Calibration curve construction
- --- Concentration calculation

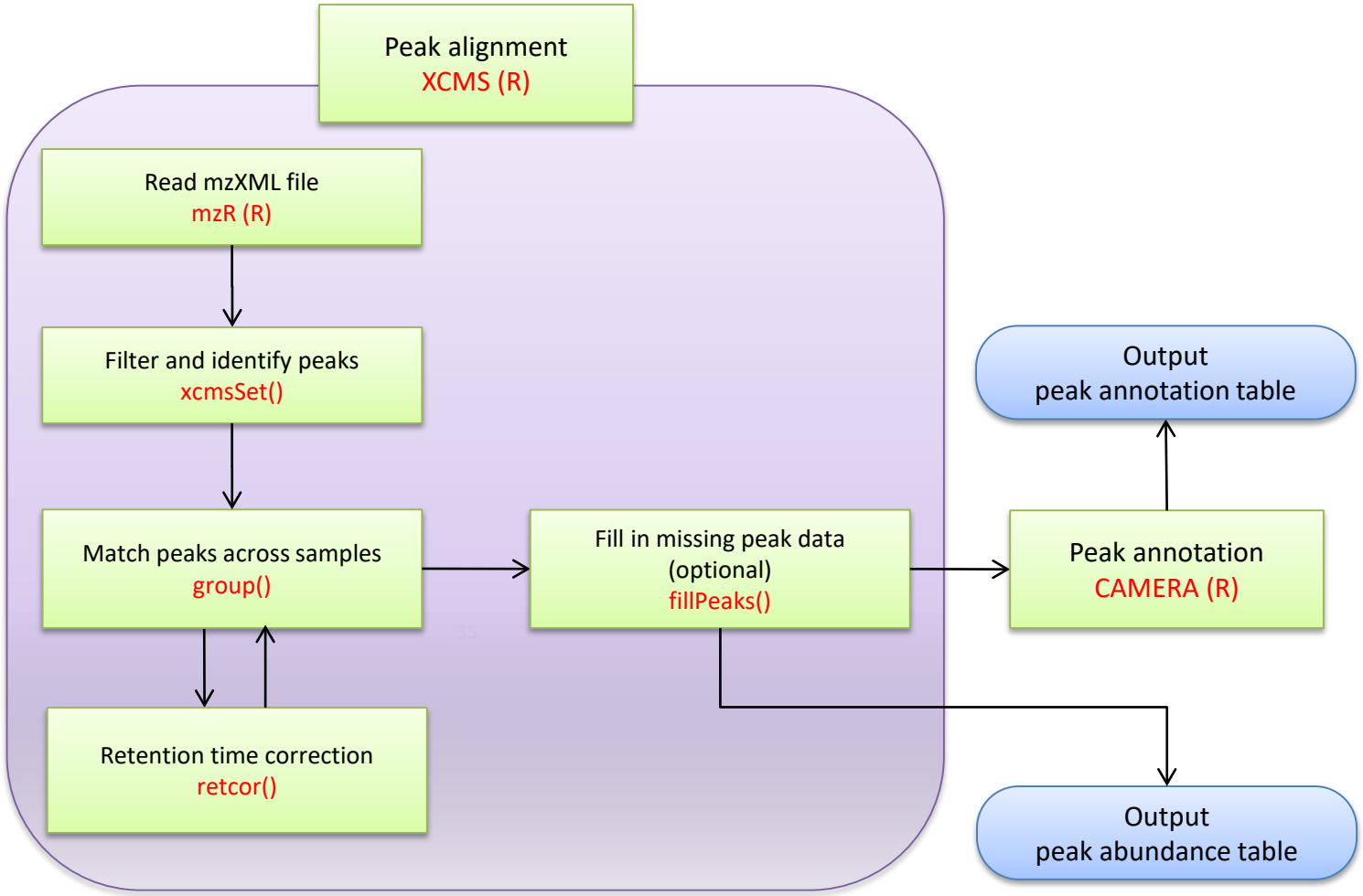
Workflow

SMART
(Statistical Metabolomics Analysis - an R Tool)



Statistical Metabolomics Analysis - an R Tool

Peak analysis (Cont.)
(Untargeted peak alignment and annotation)



Peak abundance table format:

() integrated area of original (raw) peak, () integrated area of filtered peak,

() maximum intensity of original (raw) peak, () maximum intensity of filtered peak

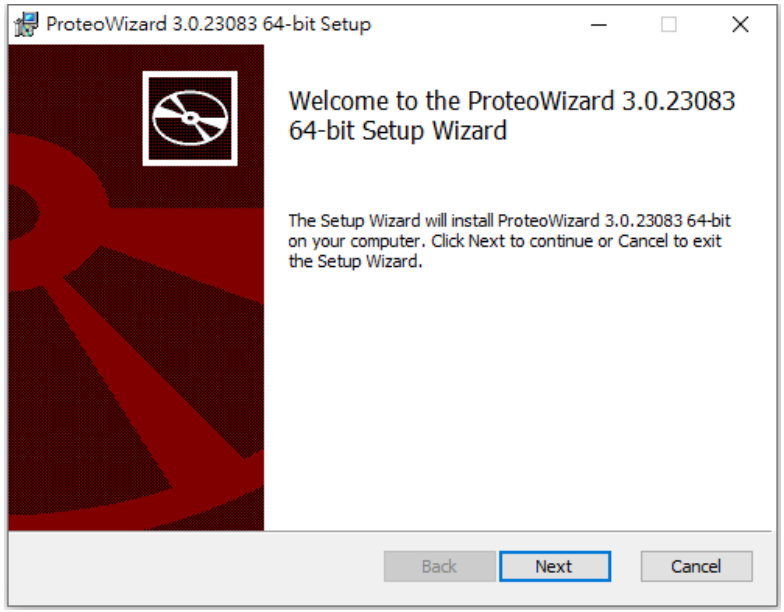
R, ActiveTCL, and ProteoWizard

SMART

SMART INSTALLATION

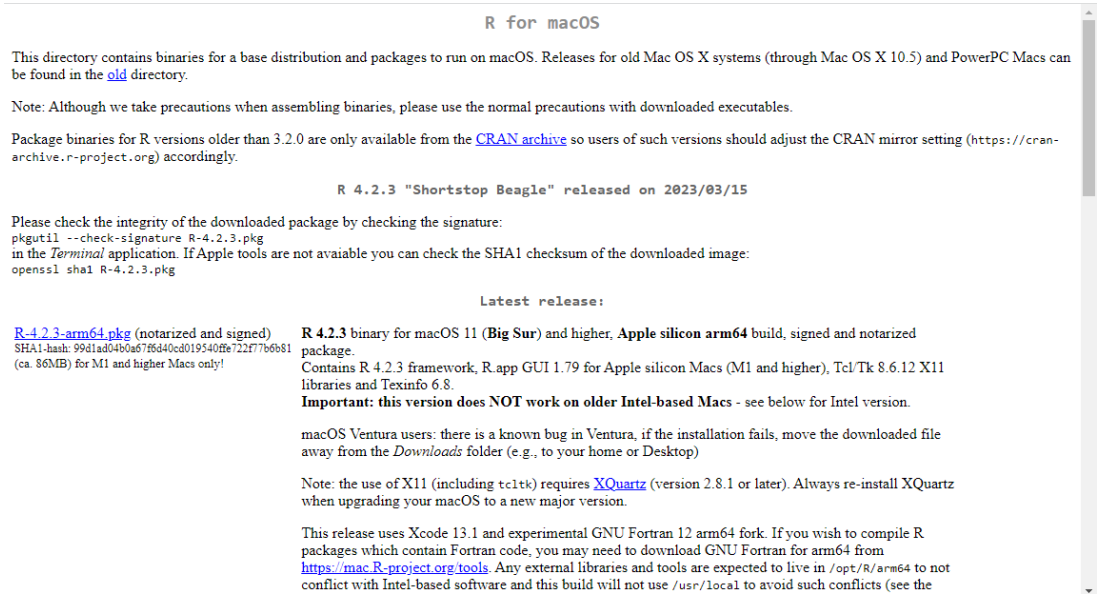
SMART installation

- Software requirements
  - **R 4.2.1** (for SMART 2.0)
    - https://cran.r-project.org/bin/windows/base/old/4.2.1/
    - https://cran.r-project.org/bin/macosx/
  - **ActiveTcl 8.6**
    - https://www.activestate.com/products/tcl/
  - **ProteoWizard 3.0**
    - https://proteowizard.sourceforge.io/download.html



SMART installation

- Mac OS (1)
  - R 4.2.X



R for macOS

This directory contains binaries for a base distribution and packages to run on macOS. Releases for old Mac OS X systems (through Mac OS X 10.5) and PowerPC Macs can be found in the old directory.

Note: Although we take precautions when assembling binaries, please use the normal precautions with downloaded executables.

Package binaries for R versions older than 3.2.0 are only available from the CRAN archive so users of such versions should adjust the CRAN mirror setting (`https://cran-archive.r-project.org`) accordingly.

R 4.2.3 "Shortstop Beagle" released on 2023/03/15

Please check the integrity of the downloaded package by checking the signature:
`pkgutil --check-signature R-4.2.3.pkg`
in the *Terminal* application. If Apple tools are not avaiable you can check the SHA1 checksum of the downloaded image:
`openssl sha1 R-4.2.3.pkg`

Latest release:

R-4.2.3-arm64.pkg (notarized and signed)
SHA1-hash: 99d1ad04b0a67f6d40cd019540ffe722f77b6b81
(ca. 86MB) for M1 and higher Macs only!

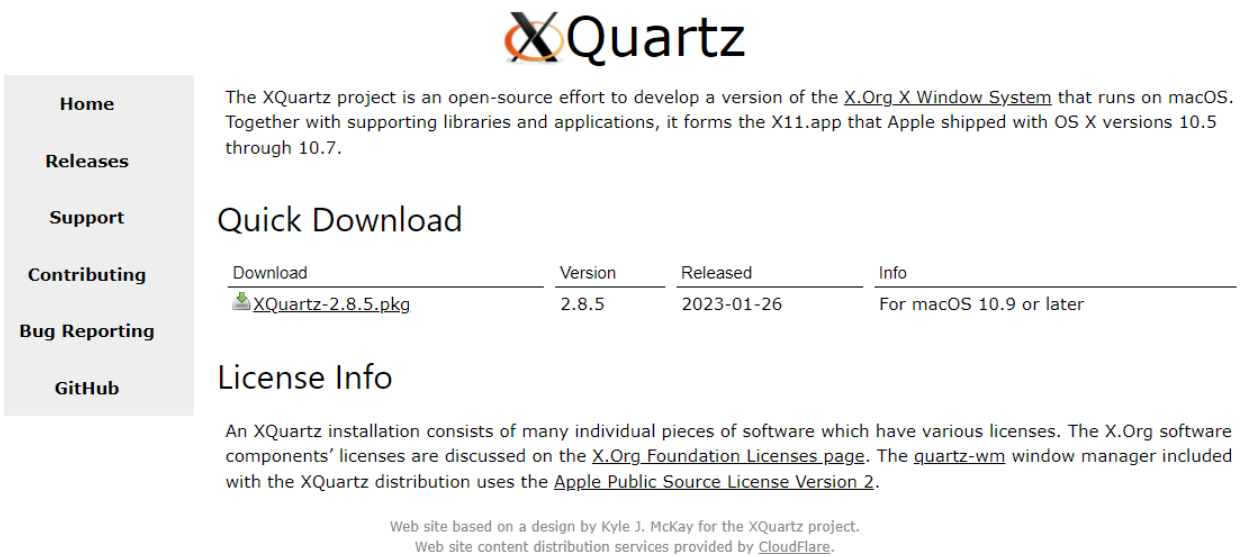
R 4.2.3 binary for macOS 11 (**Big Sur**) and higher, **Apple silicon arm64** build, signed and notarized package.
Contains R 4.2.3 framework, R.app GUI 1.79 for Apple silicon Macs (M1 and higher), Tcl/Tk 8.6.12 X11 libraries and Texinfo 6.8.
Important: this version does NOT work on older Intel-based Macs - see below for Intel version.

macOS Ventura users: there is a known bug in Ventura, if the installation fails, move the downloaded file away from the *Downloads* folder (e.g., to your home or Desktop)

Note: the use of X11 (including `tcltk`) requires XQuartz (version 2.8.1 or later). Always re-install XQuartz when upgrading your macOS to a new major version.

This release uses Xcode 13.1 and experimental GNU Fortran 12 arm64 fork. If you wish to compile R packages which contain Fortran code, you may need to download GNU Fortran for arm64 from https://mac.R-project.org/tools. Any external libraries and tools are expected to live in `/opt/R/arm64` to not conflict with Intel-based software and this build will not use `/usr/local` to avoid such conflicts (see the

  - XQuartz 2.8.5



XQuartz

Home
Releases
Support
Contributing
Bug Reporting
GitHub

The XQuartz project is an open-source effort to develop a version of the X.Org X Window System that runs on macOS. Together with supporting libraries and applications, it forms the X11.app that Apple shipped with OS X versions 10.5 through 10.7.

Quick Download

Download	Version	Released	Info
XQuartz-2.8.5.pkg	2.8.5	2023-01-26	For macOS 10.9 or later

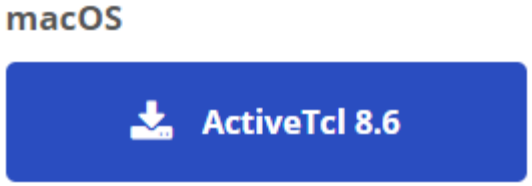
License Info

An XQuartz installation consists of many individual pieces of software which have various licenses. The X.Org software components' licenses are discussed on the X.Org Foundation Licenses page. The quartz-wm window manager included with the XQuartz distribution uses the Apple Public Source License Version 2.

Web site based on a design by Kyle J. McKay for the XQuartz project.
Web site content distribution services provided by CloudFlare.

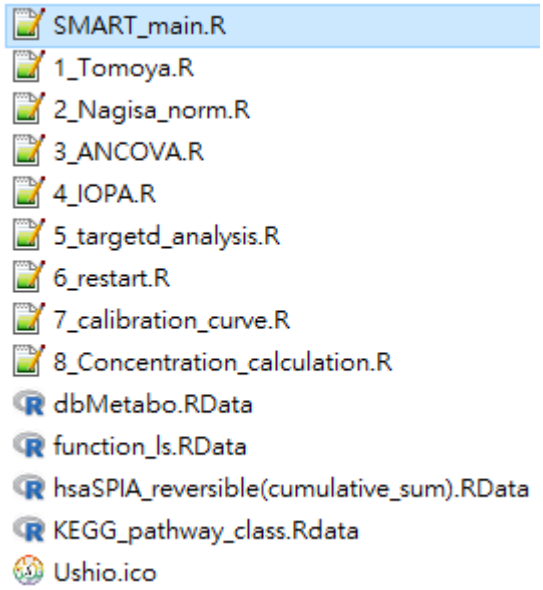
SMART installation

- Mac OS (2)
  - ActiveTcl 8.6



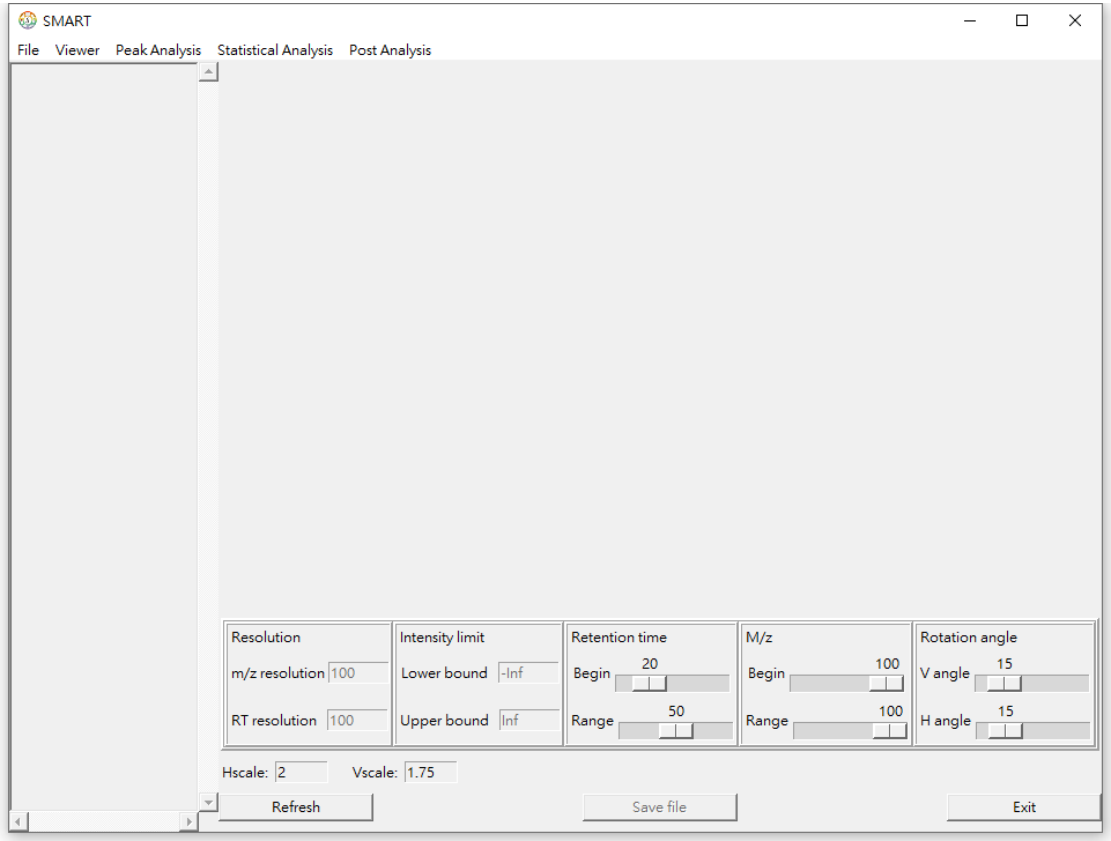
SMART execution

- Download SMART 1.2
  - http://www.stat.sinica.edu.tw/hsinchou/metabolomics/SMART.htm
  - Search by Google: "SMART metabolomics"
- Execute SMART 2.0
  - R 4.2.1
  - RStudio



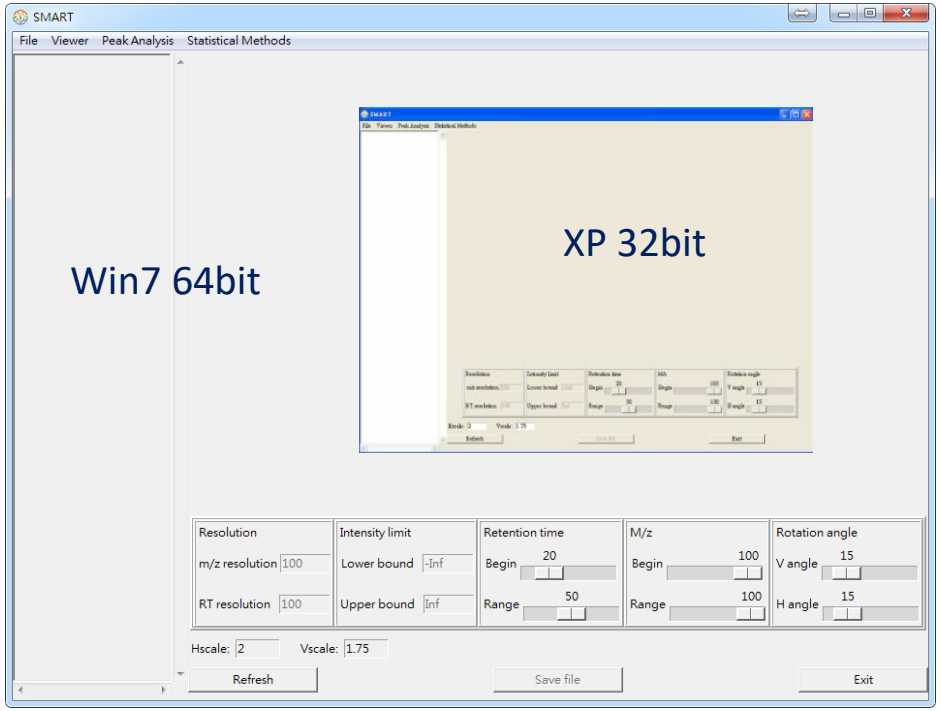
SMART_main.R

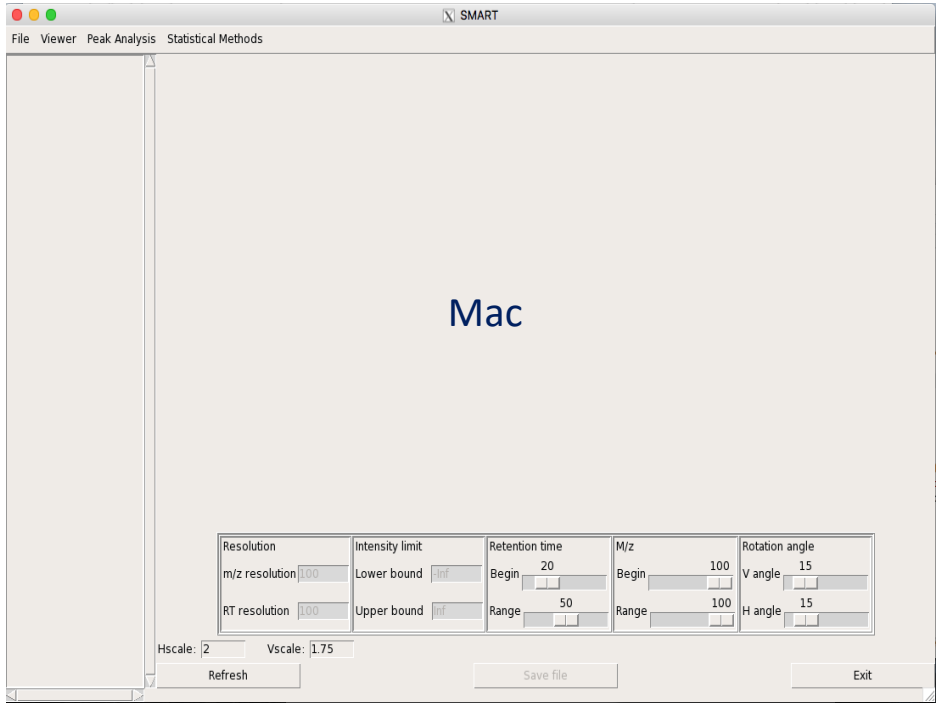
```
rm(list = ls())
gc()
library(tcltk)
setwd("G:/SMART2/Code") ##need modified##
file <- list.files(pattern = ".R")
file <- file[!grepl("RData|source",file)]
sapply(file, function(x)source(x))
load("dbMetabo.RData")
```



SMART execution

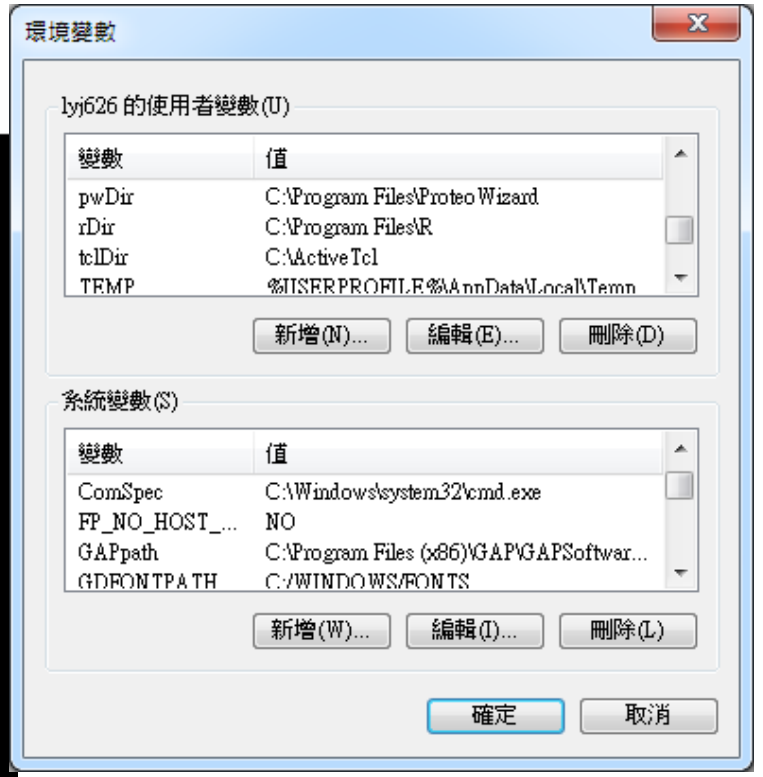
- Execute SMART (v1.1 or 1.2)
  - Windows OS (Win10, Win7, XP)
    - SMART_V1.2_64bit.exe
    - SMART_V1.1_32bit.exe
  - Mac OS
    - SMART_V1.1_Mac.r, SMART_V1.1_Mac_Gui.r, SMART_V1.1_Mac_Sub.r
    - source('~/Documents/SMART/SMART_V1.1_Mac.r')





SMART execution

- Windows (details)
  - Execute SMART*.exe



Log file

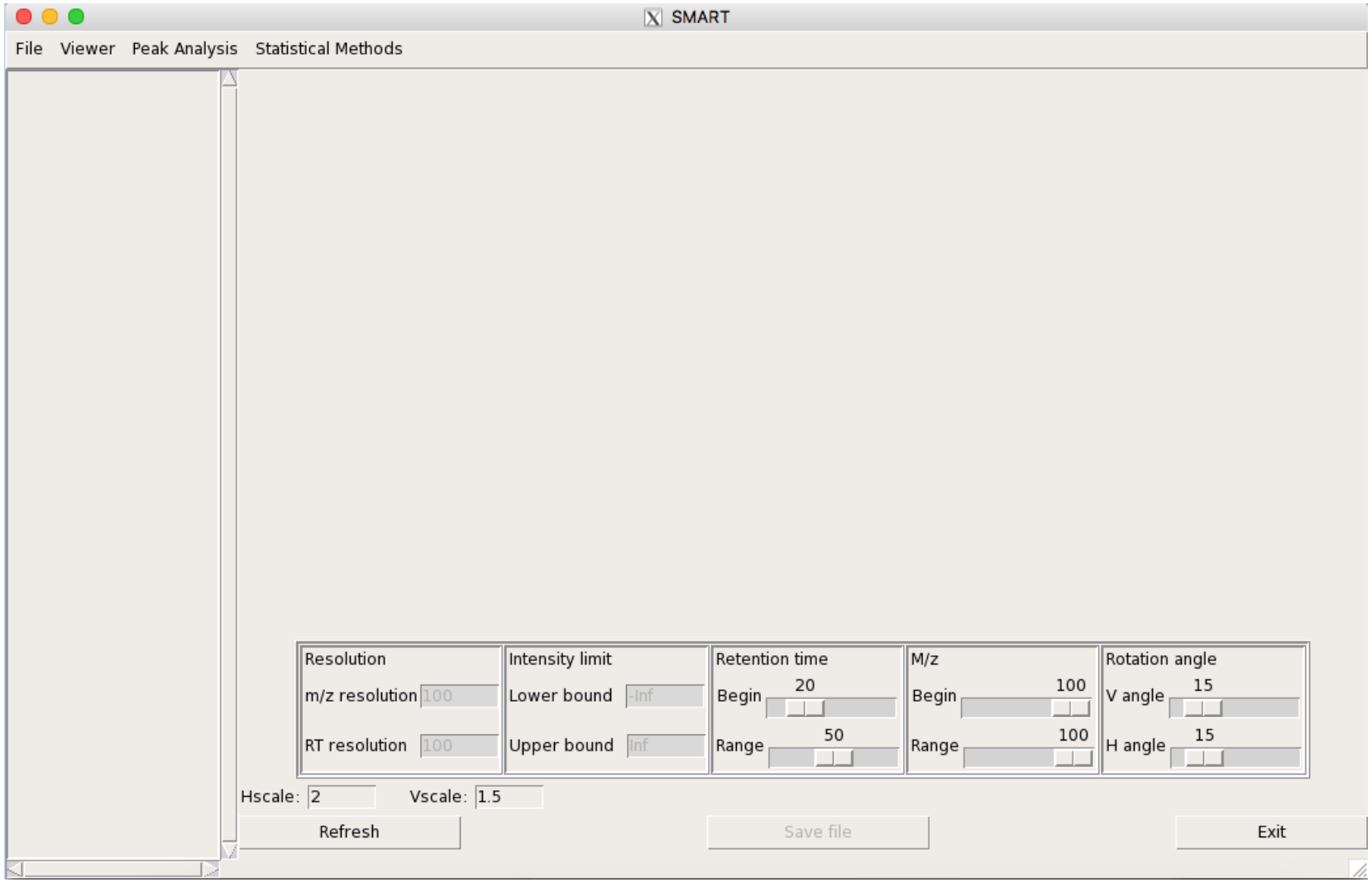
OS: Windows 7 x64
R: R version 3.3.3 (2017-03-06) x86_64
Tcl: C:\ActiveTcl

Ping cran.wu-wien.ac.at [137.208.57.37] (使用 32 位元組的資料)
回覆自 137.208.57.37: 位元組=32 時間=294ms TTL=43

137.208.57.37 的 Ping 統計資料:
封包: 已傳送 = 1，已收到 = 1, 已遺失 = 0 (0% 遺失)，

SMART execution

- GUI



HT (hypertension), Drug, and BC (breast cancer)

DEMONSTRATION DATA

Demonstration data (HT)

Dr. Pan, Wen-Harn 潘文涵老師 – 生醫所　　LC-MS

- Pharmacometabolomics study (Comput Struct Biotechnol J. 2022)
  - –250 hypertension patients
  - –Select 10 hypertension patients
    - 5 ACEi treated patients
    - 5 non-treated patients
    - 60 replicate samples

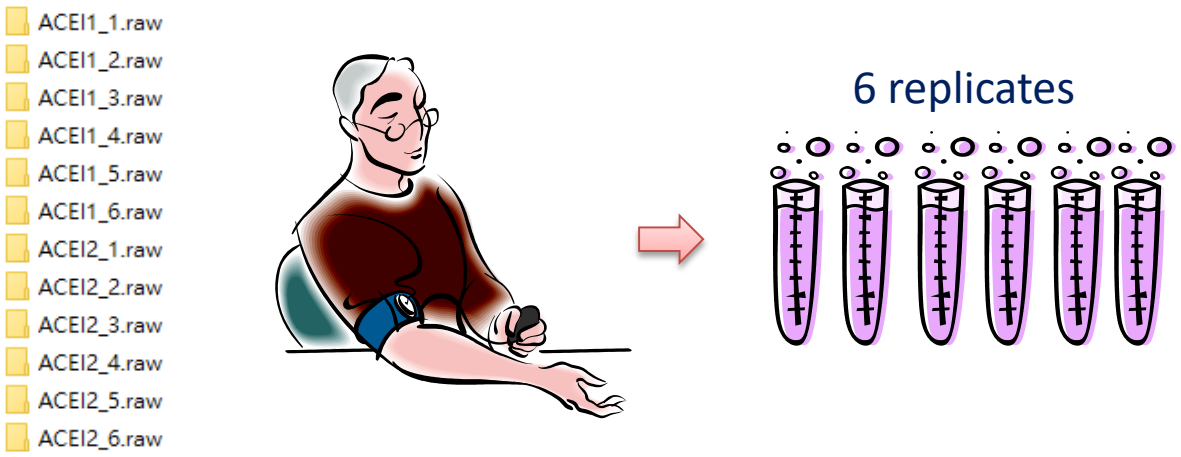
ELSEVIER
Guide for Authors | About | Explore this Journal
COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

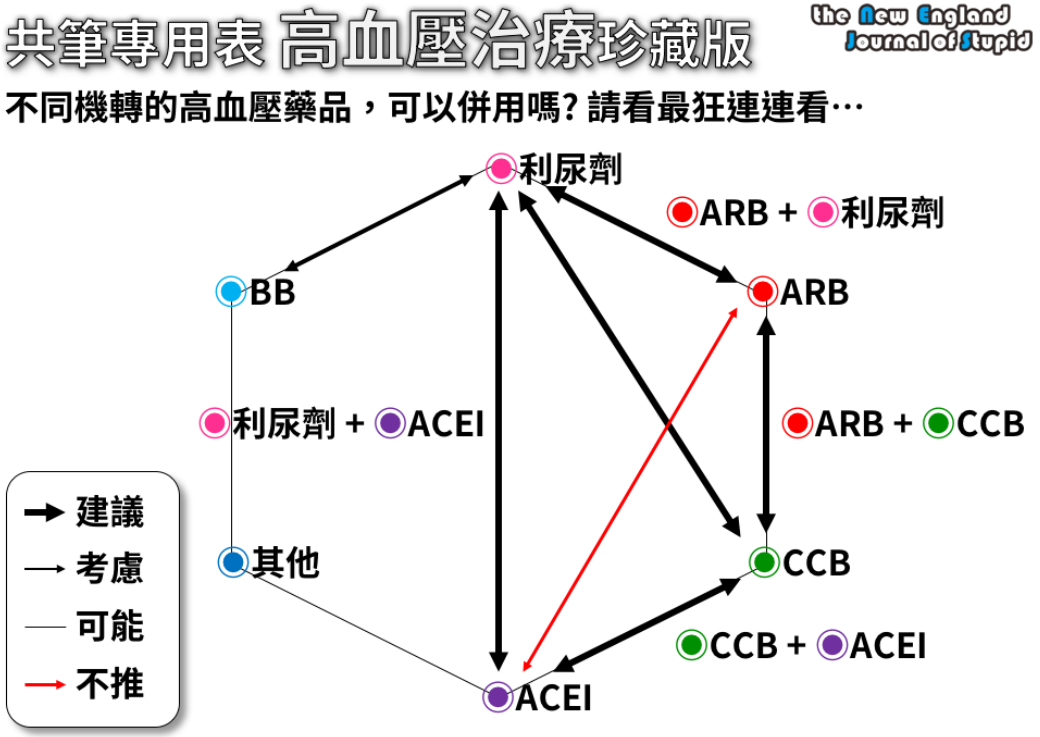
Comput Struct Biotechnol J. 2022; 20: 6458–6466.
Published online 2022 Nov 17. doi: 10.1016/j.csbj.2022.11.030
PMCID: PMC9708446
PMID: 36467587

Pharmacometabolomic study of drug response to antihypertensive medications for hypertension marker identification in Han Chinese individuals in Taiwan

Yu-Jen Liang,[a,1] Kuang-Mao Chiang,[b,1] Li-li Xiu,[c] Chia-Min Chung,[d] Chi-Jen Lo,[e,f] Ming-Shi Shiao,[e,f] Mei-Ling Cheng,[e,f] Cheng-Chin Kuo,[c] Hsin-Chou Yang,[a,g,h,i,*] and Wen-Harn Pan[b,*]

Comput Struct Biotechnol J





Demonstration data (Drug)

LC-MS/MS

Dr. Wang, Yi Sheng 王亦生老師 - 基因體研究中心

	Drug Name
1	海洛英 (Heroin) (Diacetylmorphine)
2	嗎啡 (Morphine)
3	古柯鹼 (Cocaine)
4	Thebaine
5	大麻 (delta9-THC) (delta9-Tetrahydrocannabinol)
6	安非他命 (Amphetamine)
7	甲基安非他命 (MA) (Methamphetamine)
8	搖頭丸 (MDMA) (3,4-Methylenedioxymethamphetamine)
9	Love Drug或Mellow Drug of America MDA (3,4-Methylenedioxyamphetamine)
10	愷他命 (Ketamine)
11	氟硝西泮（Rohypnol）Flunitrazepam (FM2)
12	硝甲西泮（硝甲氮平）Nimetazepam

.raw file	
12種毒品 (11 standard samples) 序列式稀釋混合	(1) 50 ppb
	(2) 100 ppb
	(3) 200 ppb
D1_50_MS1.mzXML D2_100_MS1.mzXML D3_200_MS1.mzXML D4_300_MS1.mzXML D5_400_MS1.mzXML D6_500_MS1.mzXML D7_600_MS1.mzXML D8_700_MS1.mzXML D9_800_MS1.mzXML D10_900_MS1.mzXML D11_1000_MS1.mzXML	D1_50_MSMS.mzXML D2_100_MSMS.mzXML D3_200_MSMS.mzXML D4_300_MSMS.mzXML D5_400_MSMS.mzXML D6_500_MSMS.mzXML D7_600_MSMS.mzXML D8_700_MSMS.mzXML D9_800_MSMS.mzXML D10_900_MSMS.mzXML D11_1000_MSMS.mzXML
	(4) 300 ppb
	(5) 400 ppb
	(6) 500 ppb (for test)
	(7) 600 ppb
	(8) 700 ppb
	(9) 800 ppb
	(10) 900 ppb
	(11) 1000 ppb

Demonstration data (Drug)

LC-MS/MS

- Parameter setting : the tolerance of m/z and retention time (RT)
- Input file : the true m/z information of m/z (required), RT (optional) and RT tolerance(optional)

- Output file :
  - Peak abundance table
  - MS1 and MS2 spectrum
  - SAMPLENAME_target-detection.pdf (future)

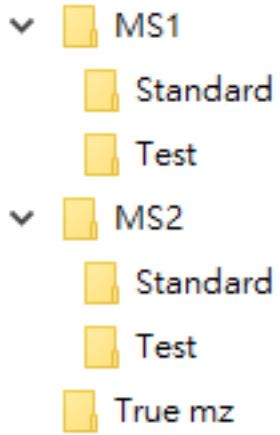
Data	Ans	FULL MS & MSMS				FULL MS only		MSMS only	
MS1	RT	v	v			v			
	mz	v	v	v	v	v	v		
MS2	RT	v			v			v	
	mz	v	v	v	v			v	v

Name	m/z	RT	RT tolerance
Heroin	370	11.42	
morphine	286	2.45	
cocaine	304	11.7	
Thebaine	312	11.31	
delta-9 THC	315	25.23	
Amphetamine	136	7.04	
MA	150	8.26	
MDMA	194	8.99	
Love Drug	180	8.44	
Ketamine	238	10	
FM2	314	17.4	
Nimetazepam	296	17.35	

Input format of true m/z information file (MS1)

Name	m/z
Heroin	370, 328, 268, 211, 193, 58
Morphine	286, 268, 229, 201
Cocaine	304, 182, 150, 105, 82
Thebaine	312, 281, 266, 251, 221, 58
delta-9 THC	315, 259, 193, 135, 123, 107, 93
Amphetamine	136, 119, 91
MA	150, 119, 91
MDMA	194, 163, 135, 133, 105
Love Drug	180, 163, 135, 133, 105
Ketamine	238, 220, 179, 163, 152, 125
FM2	314, 200, 286, 268
Nimetazepam	296, 268, 250, 222, 193, 165

Input format of true m/z information file (MS2)



Demonstration data (BC)

Journal List > J Clin Invest > v.124(1); 2014 Jan 2 > PMC3871244

JCI The Journal of Clinical Investigation
Published by The American Society for Clinical Investigation | Founded 1908

J Clin Invest. 2014 Jan 2; 124(1): 398–412.
Published online 2013 Dec 9. doi: 10.1172/JCI71180
PMCID: PMC3871244
PMID: 24316975

致癌代謝物

MYC-driven accumulation of 2-hydroxyglutarate is associated with breast cancer prognosis

Atsushi Terunuma,[1] Nagireddy Putluri,[2] Prachi Mishra,[1] Ewy A. Mathé,[1] Tiffany H. Dorsey,[1] Ming Yi,[3] Tiffany A. Wallace,[1]

- Breast cancer study

  132 samples

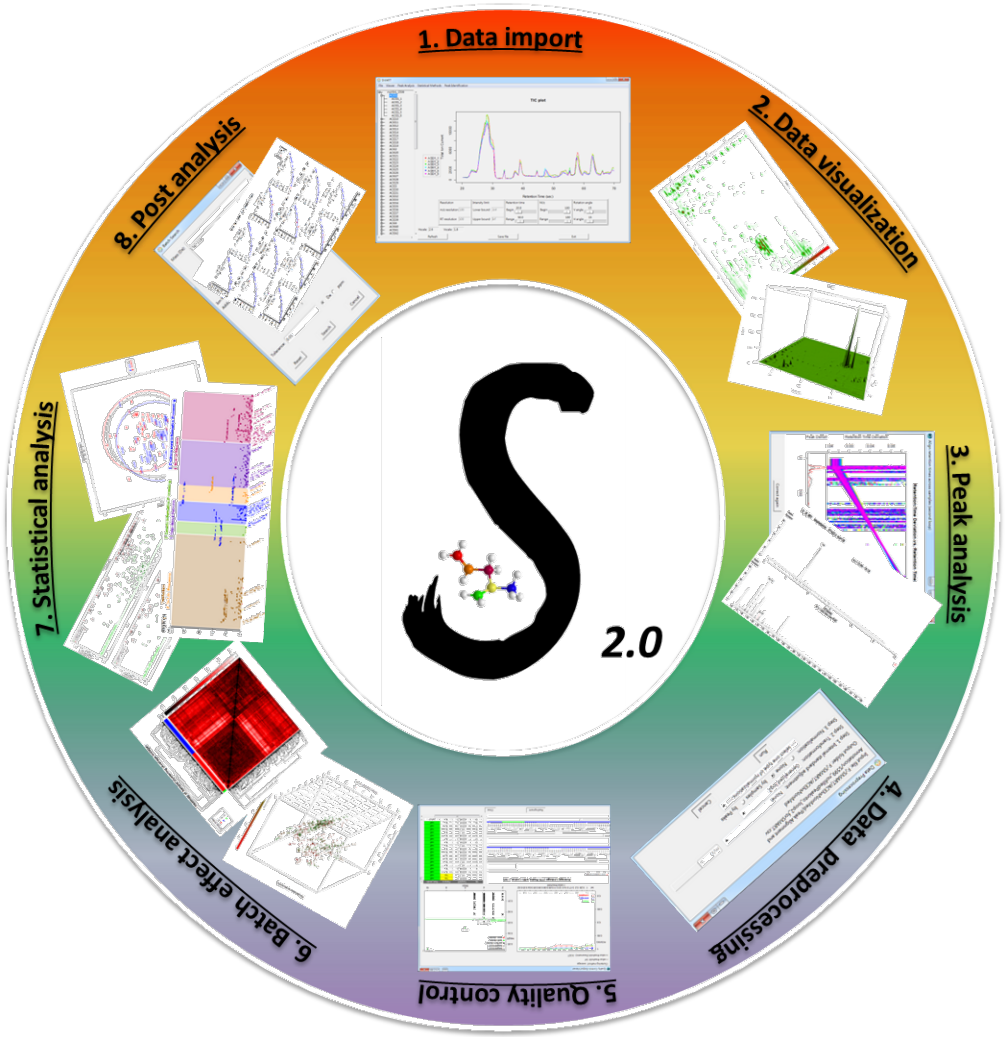
  64 Black African-Americans and 68 White European-Americans.

  –Gene expression (GEO GSE37751) Affymetrix Human Gene 1.0 ST Array
  - 108 samples. (53 B, 55 W)
    - Breast tumor samples (n = 61) and adjacent non-tumor tissue samples (n = 47).
  - 20254 genes.

  –Metabolomics (Supplemental Tables 1 and 2) GC-MS
  - 132 samples. (64 B, 68 W)
    - 67 human breast tumors and 65 tumor-adjacent noncancerous tissues.
  - 536 metabolites.
    - 32 metabolites are all zero.

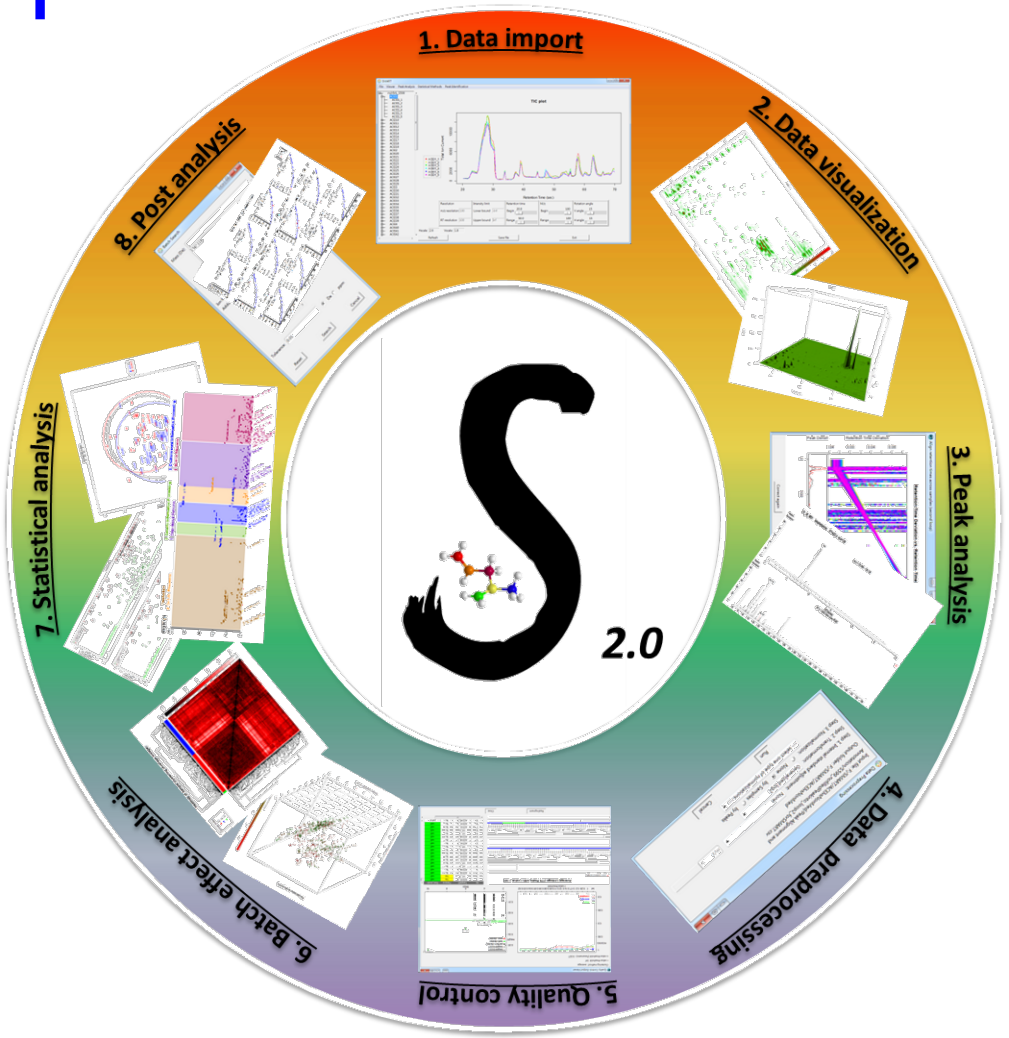
乳腺癌中，致癌代謝物 2-羥基戊二酸 (2HG) 的濃度升高與 MYC pathway之間有關聯。

SMART OPERATION

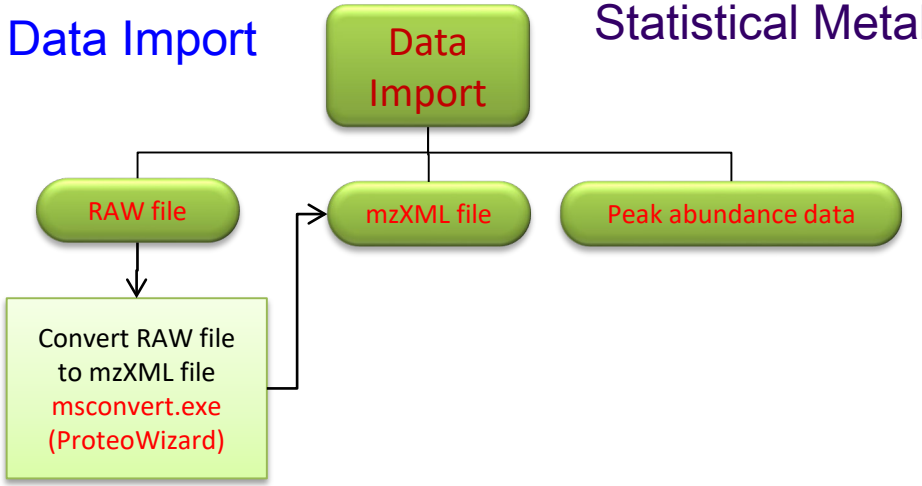


Statistical Metabolomics Analysis – an R Tool

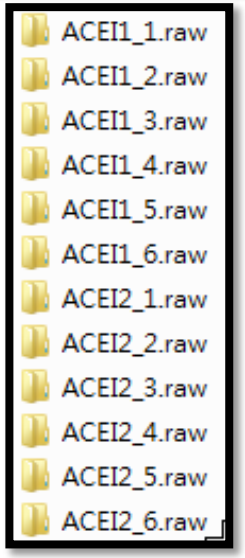
1. Data import



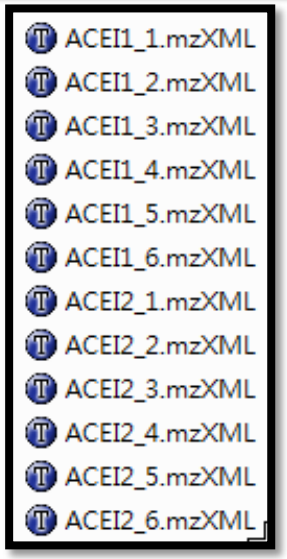
Data Import



1 RAW file (.raw or .d)



2 mzXML format file (.mzXML)



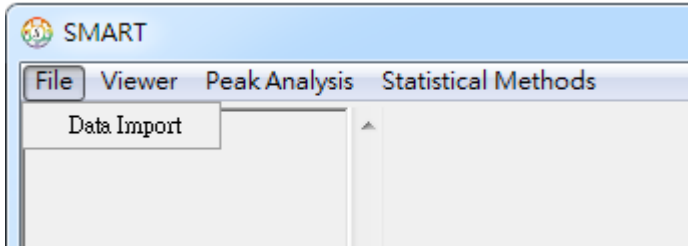
3 Peak abundance data file (.csv)

Peak_Index	mz	Ret_Time.sec.	ACEI1_1	ACEI1_2	ACEI1_3
1	55.9369	336.6774	40.48807	49.20359	73.12604
2	56.9635	336.0434	26.2133	NA	34.25459
3	60.07568	29.1904	15.04558	18.71559	15.56744
4	61.01206	39.1728	NA	NA	NA
5	68.9834	23.9603	32.57462	39.53168	40.60577
6	69.07024	53.4746	NA	NA	NA
7	70.06568	32.2958	NA	NA	NA
8	72.08126	64.0674	NA	NA	NA
9	72.08131	37.8029	75.99819	59.4558	52.61279
10	77.03941	58.097	36.12143	39.46023	38.78575

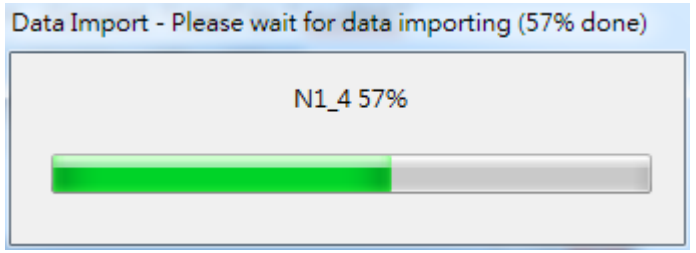
Data import (raw file)

- Data import (e.g., .raw file)
  - Input folder
    - indicate the directory of ".raw" files
  - Output folder
    - Create "mzXML" to be the directory name in the output folder
    - Convert ".raw" to ".mzXML"

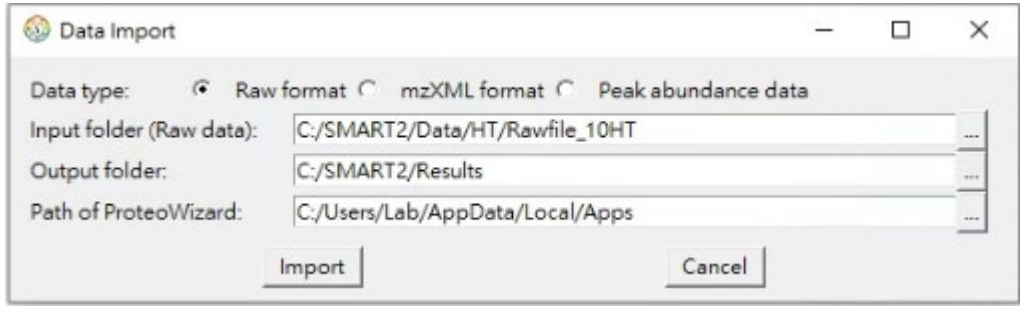
1



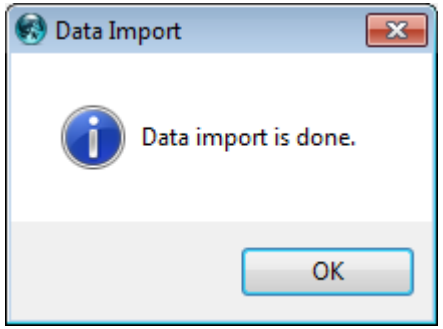
3



2

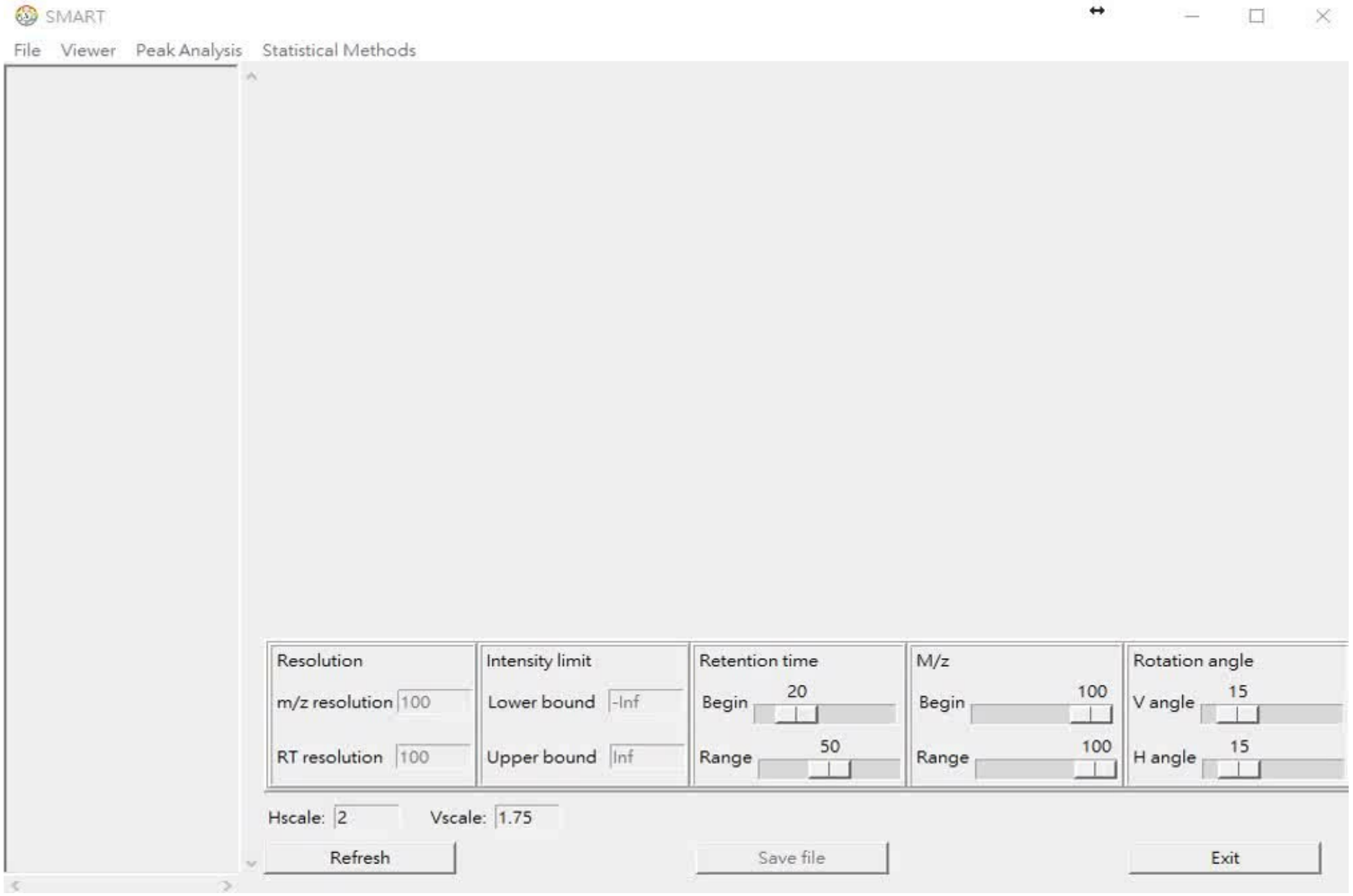


4



Note: only ".raw" file can be in the input directory

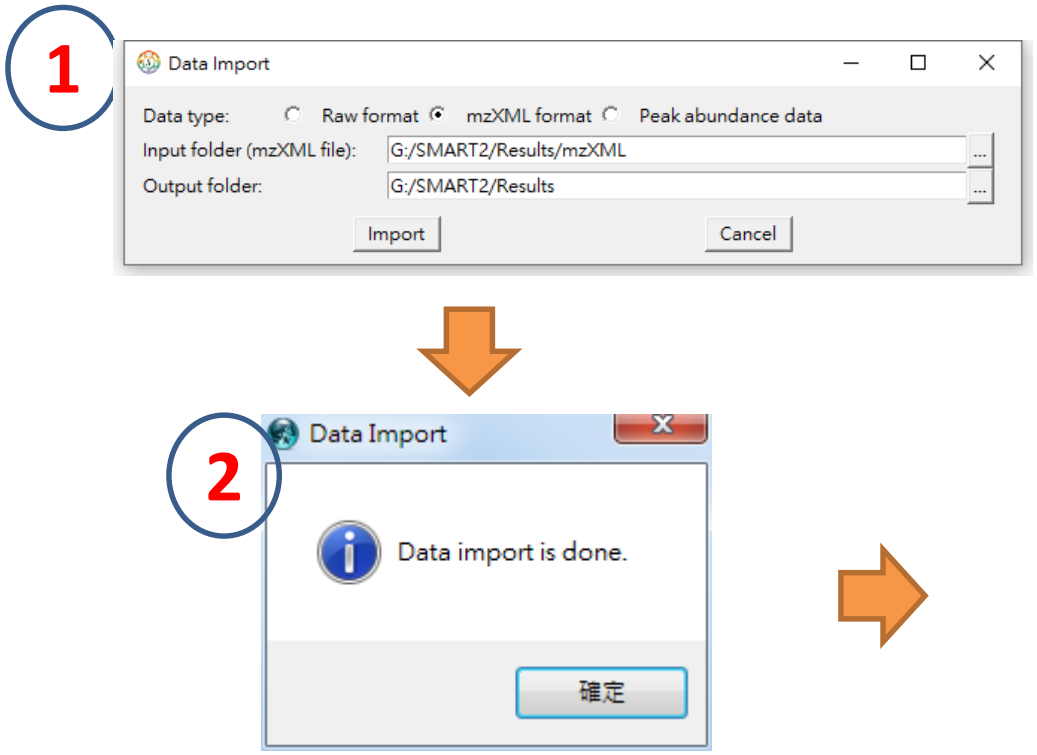
ProteoWizard:
C:\Users*username*\AppData\Local\Apps

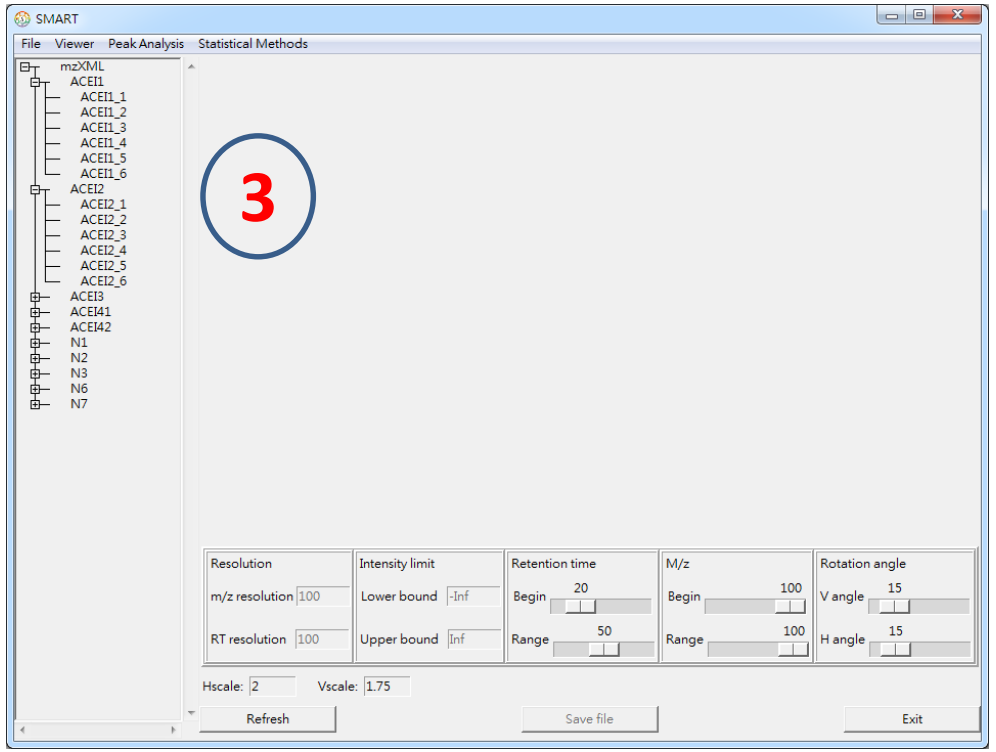


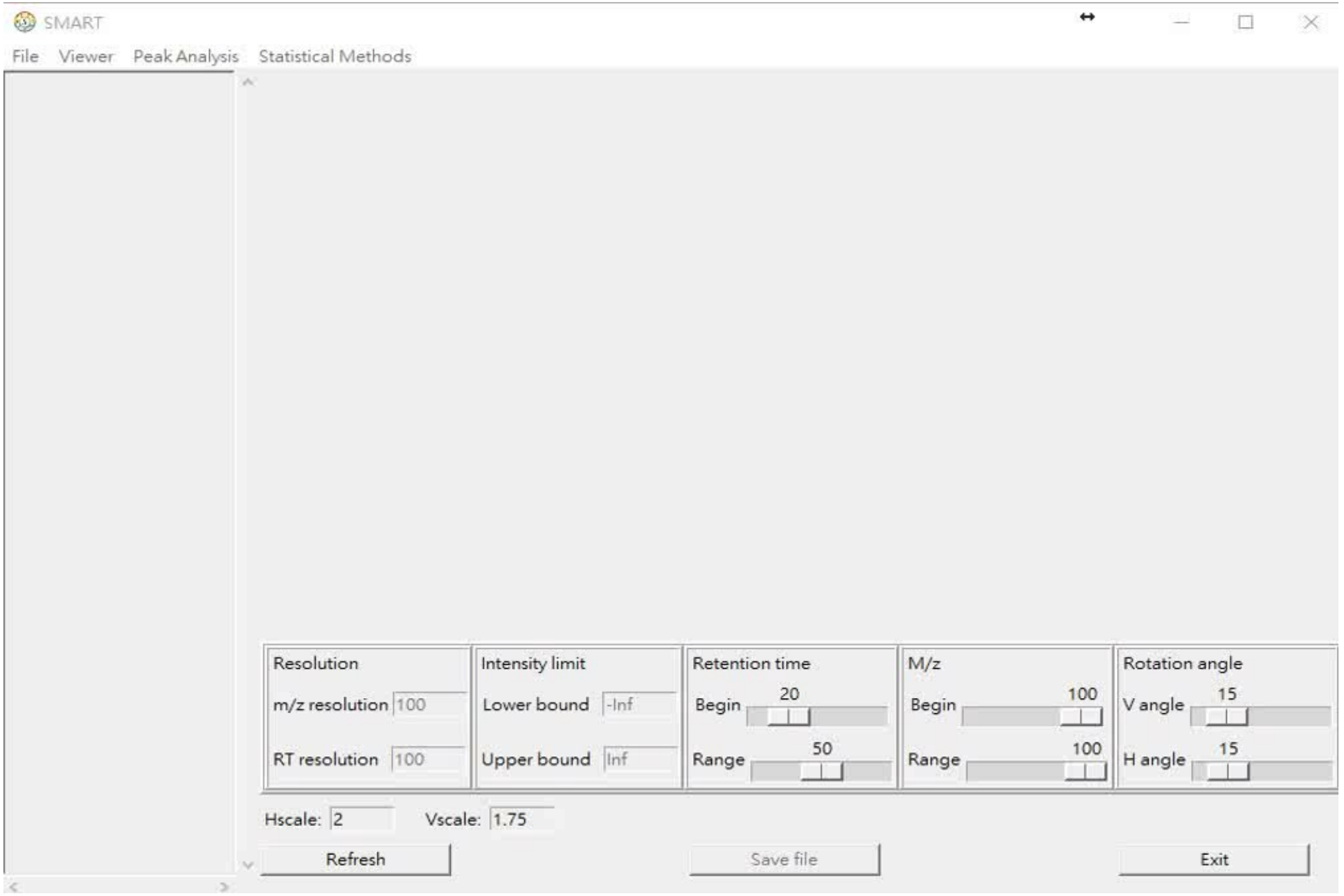
Data import (for RAW file)

Data import (mzXML file)

- Data import (for mzXML file)
  - Folder setting
    - Note: only “mzXML” file can be in the input folder



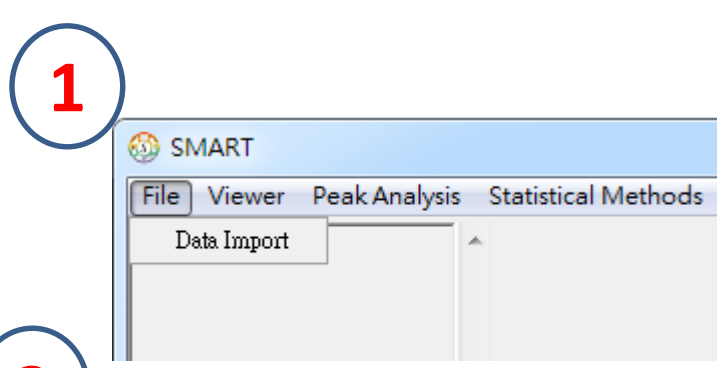


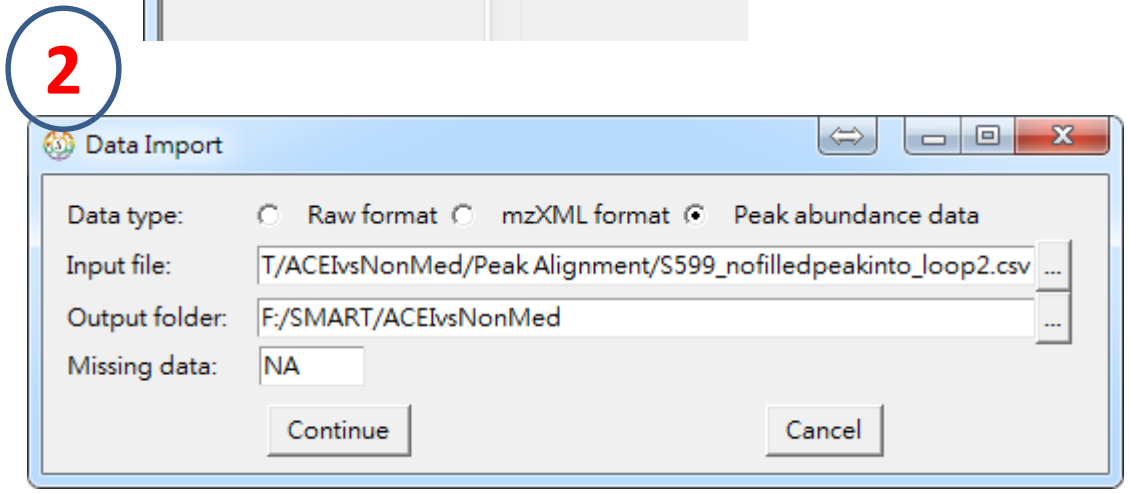


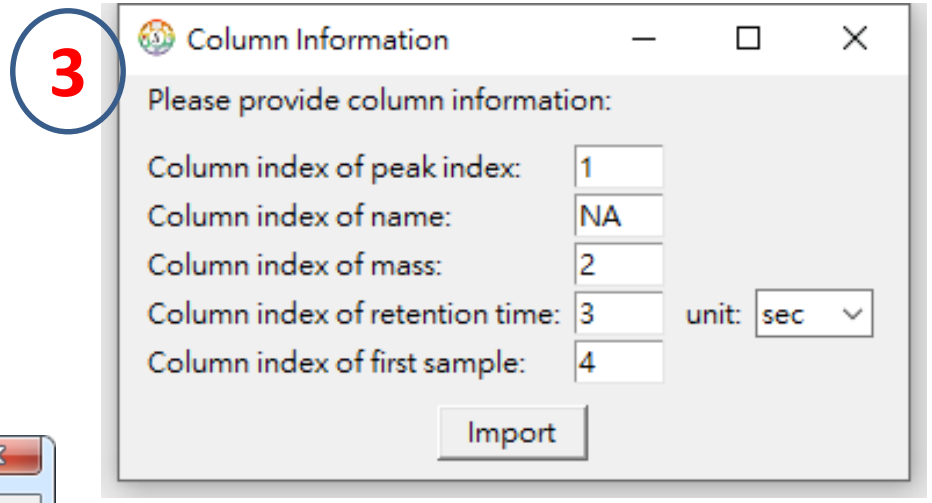
Import mzXML

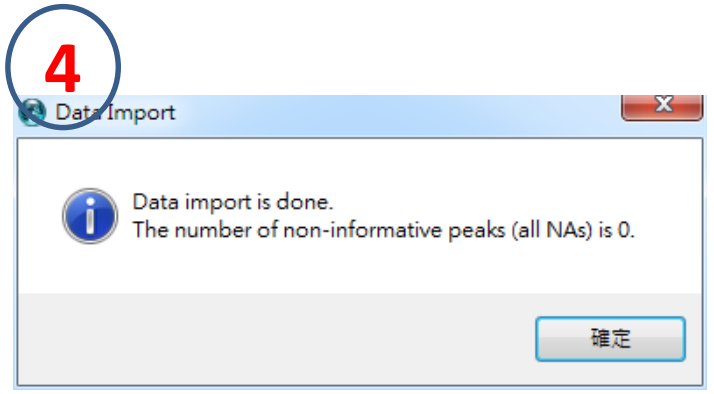
Data import (peak abundance data)

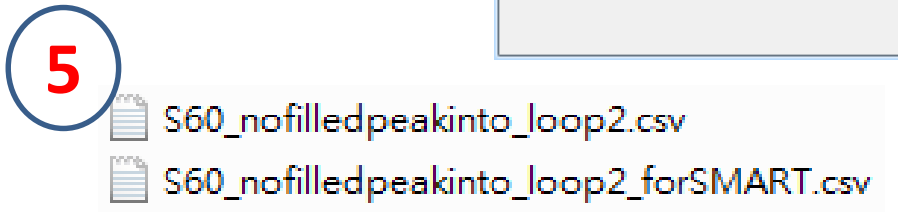
- Data import (for peak abundance data)





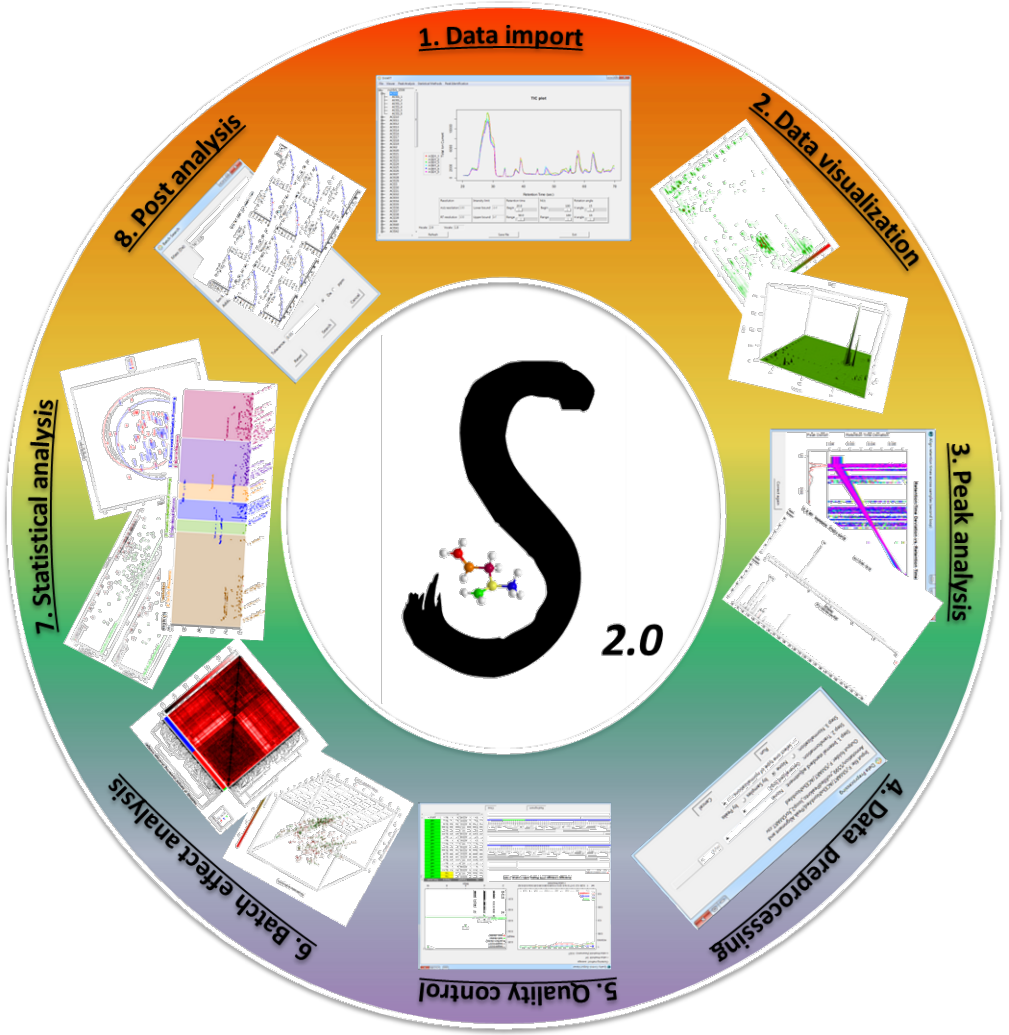






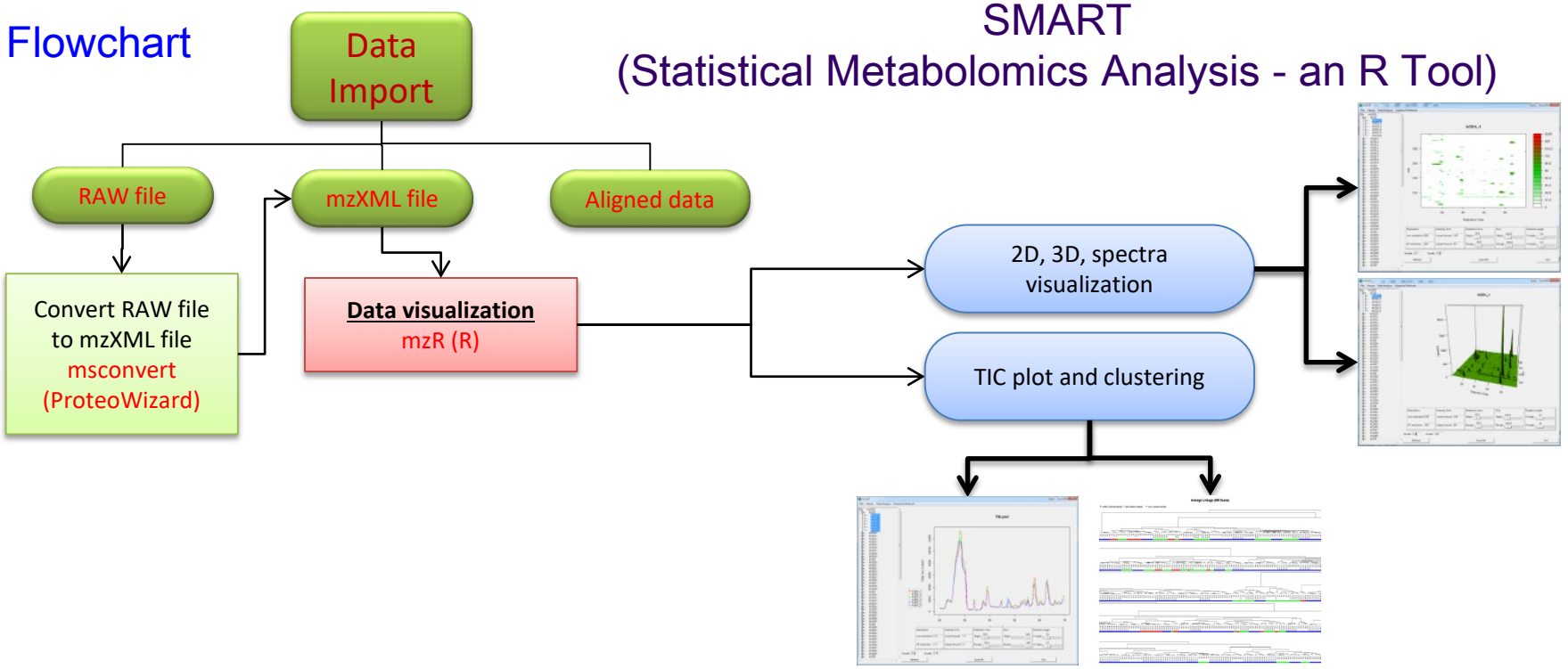
Data import (for peak abundance data)

2. Data visualization



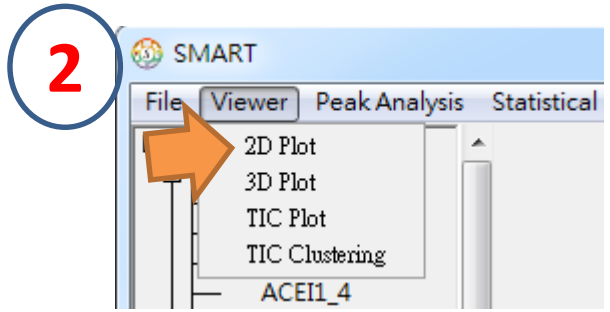
SMART
(Statistical Metabolomics Analysis - an R Tool)

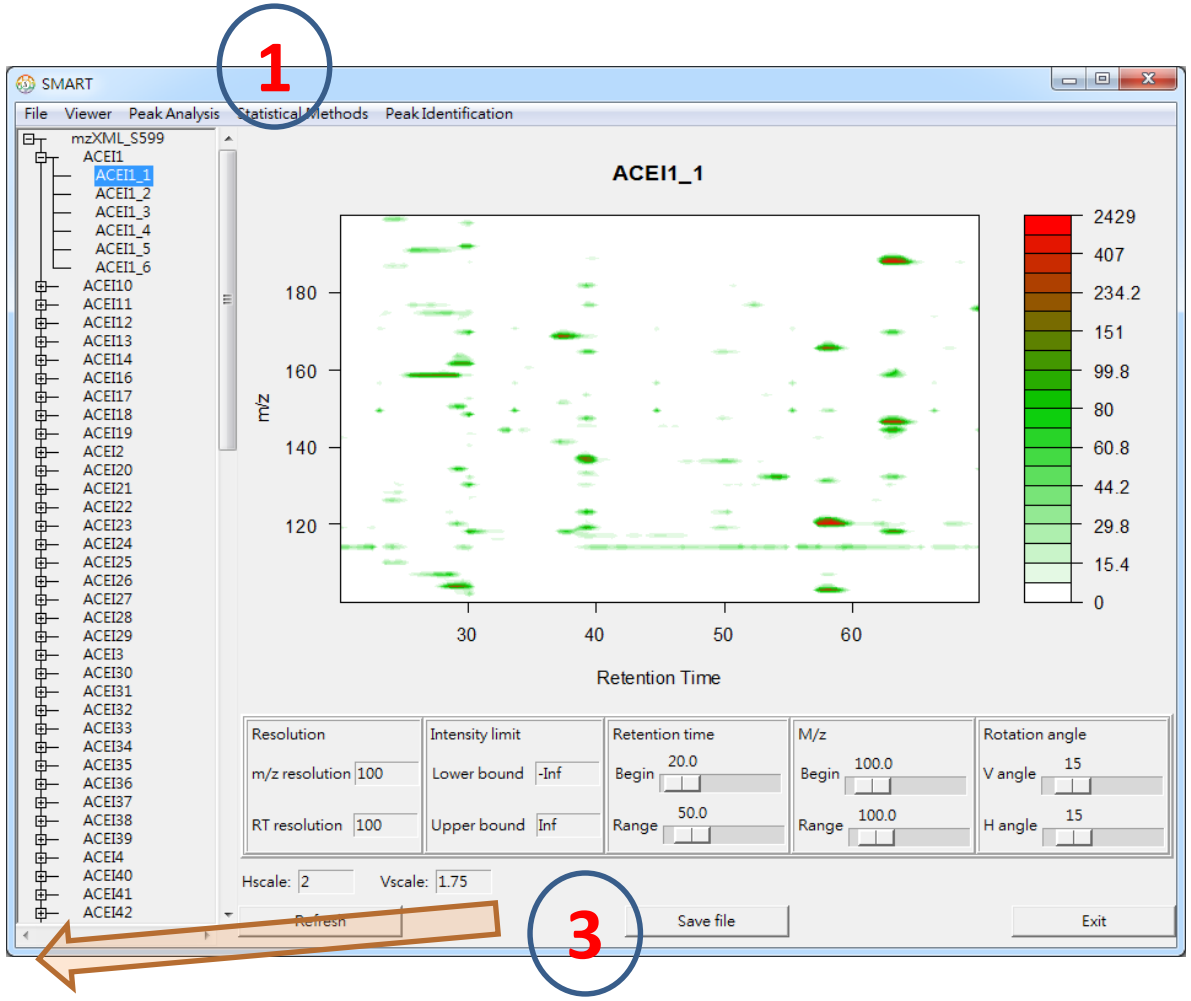
Flowchart

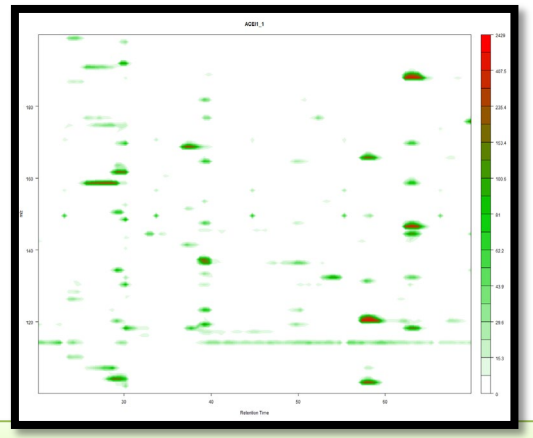


Data visualization

- 2D plot



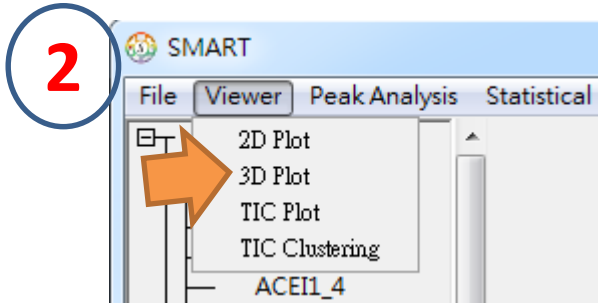


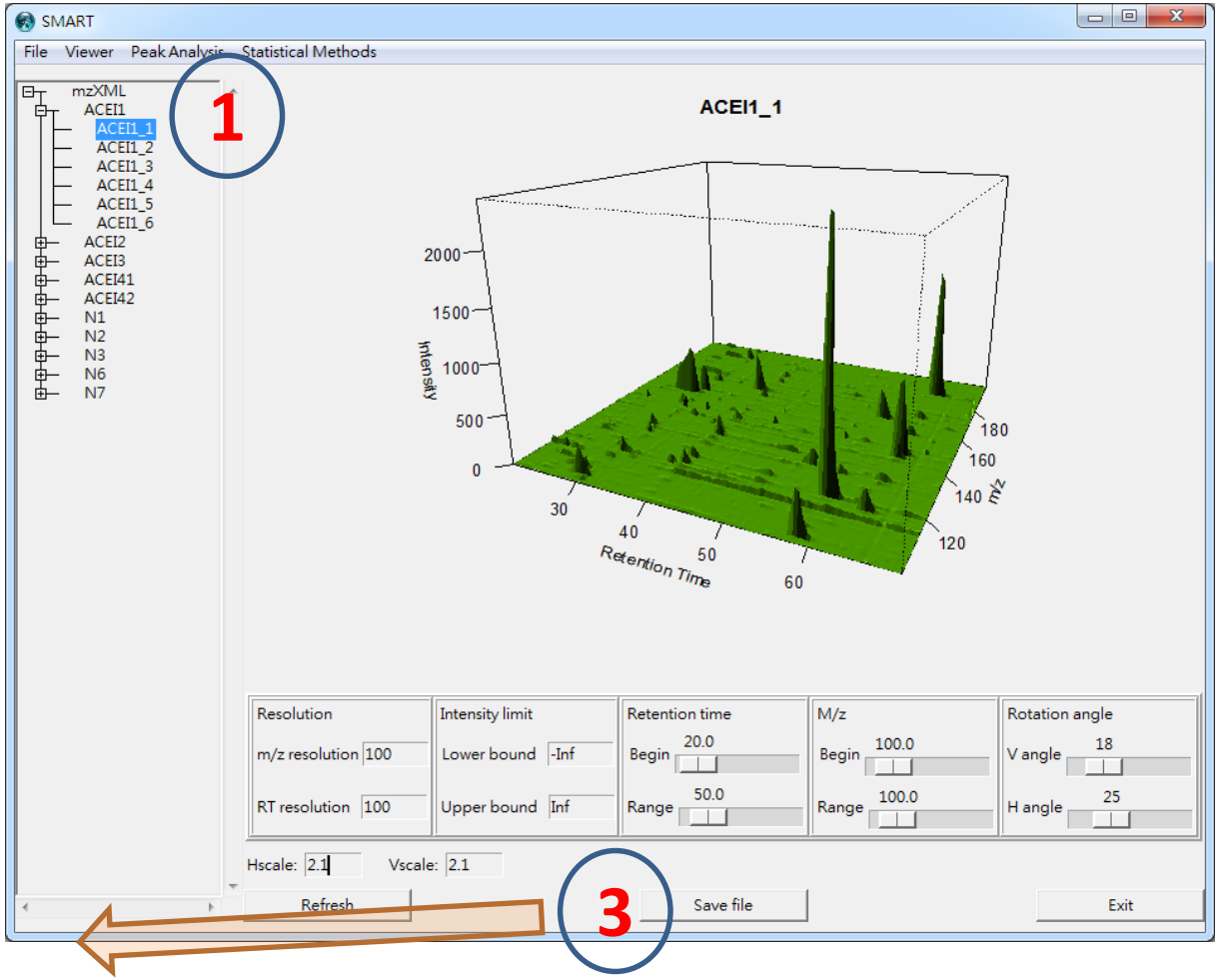


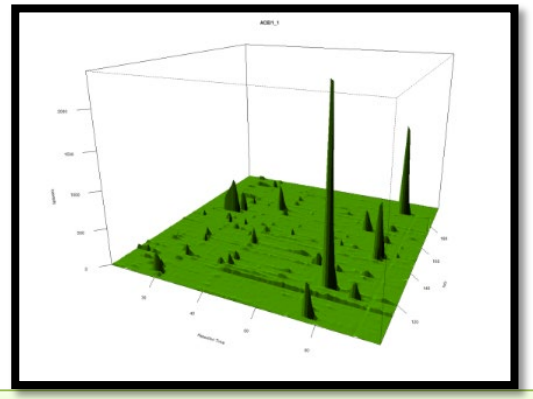
outpath\2D plot\ACEI1_1_RT20(50)_mz100(100).png

Data visualization

- 3D plot



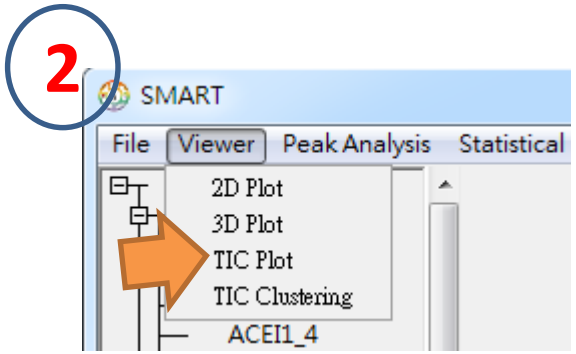


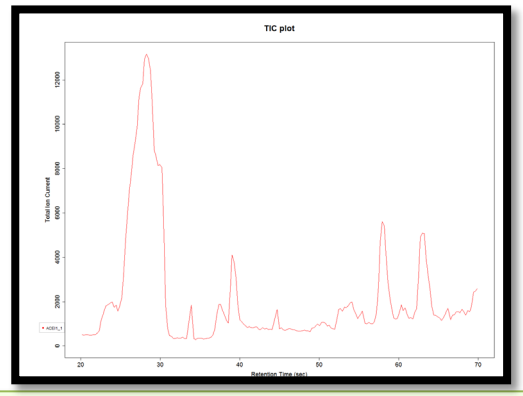


outpath\3D plot\ACEI1_1_RT20(50)_mz100(100)_h25_v18.png

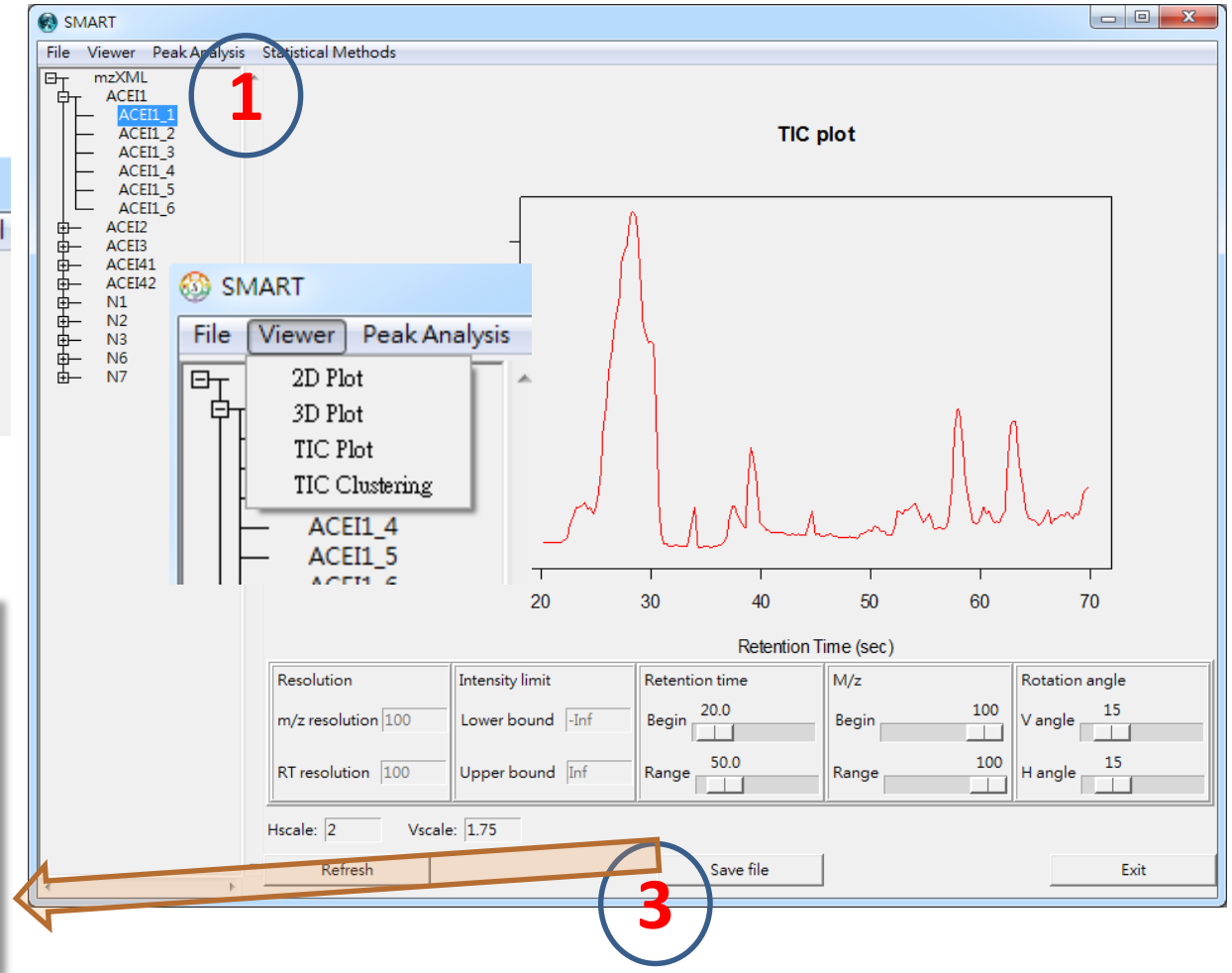
Data visualization

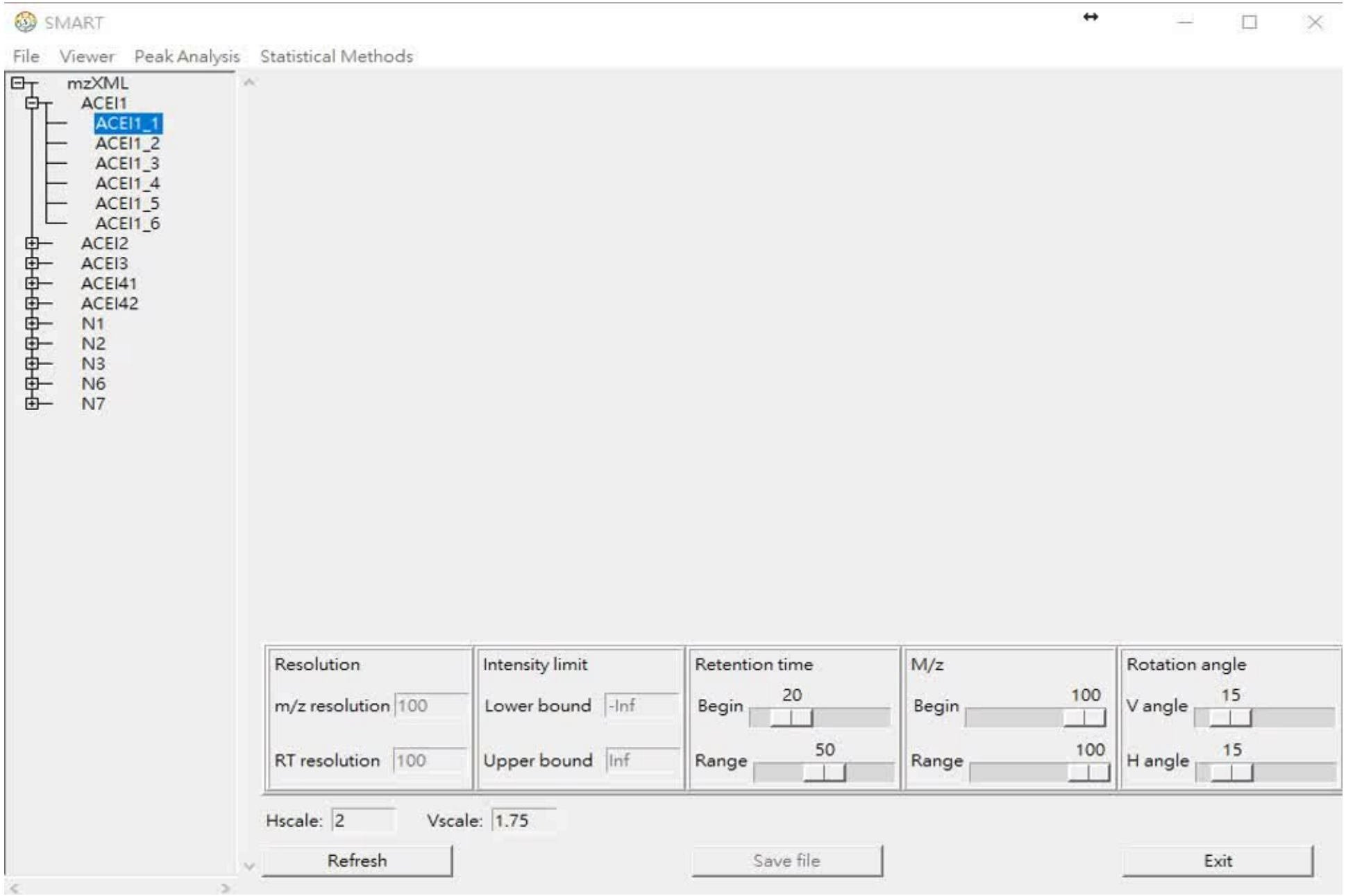
- TIC plot
  - For each replicate sample



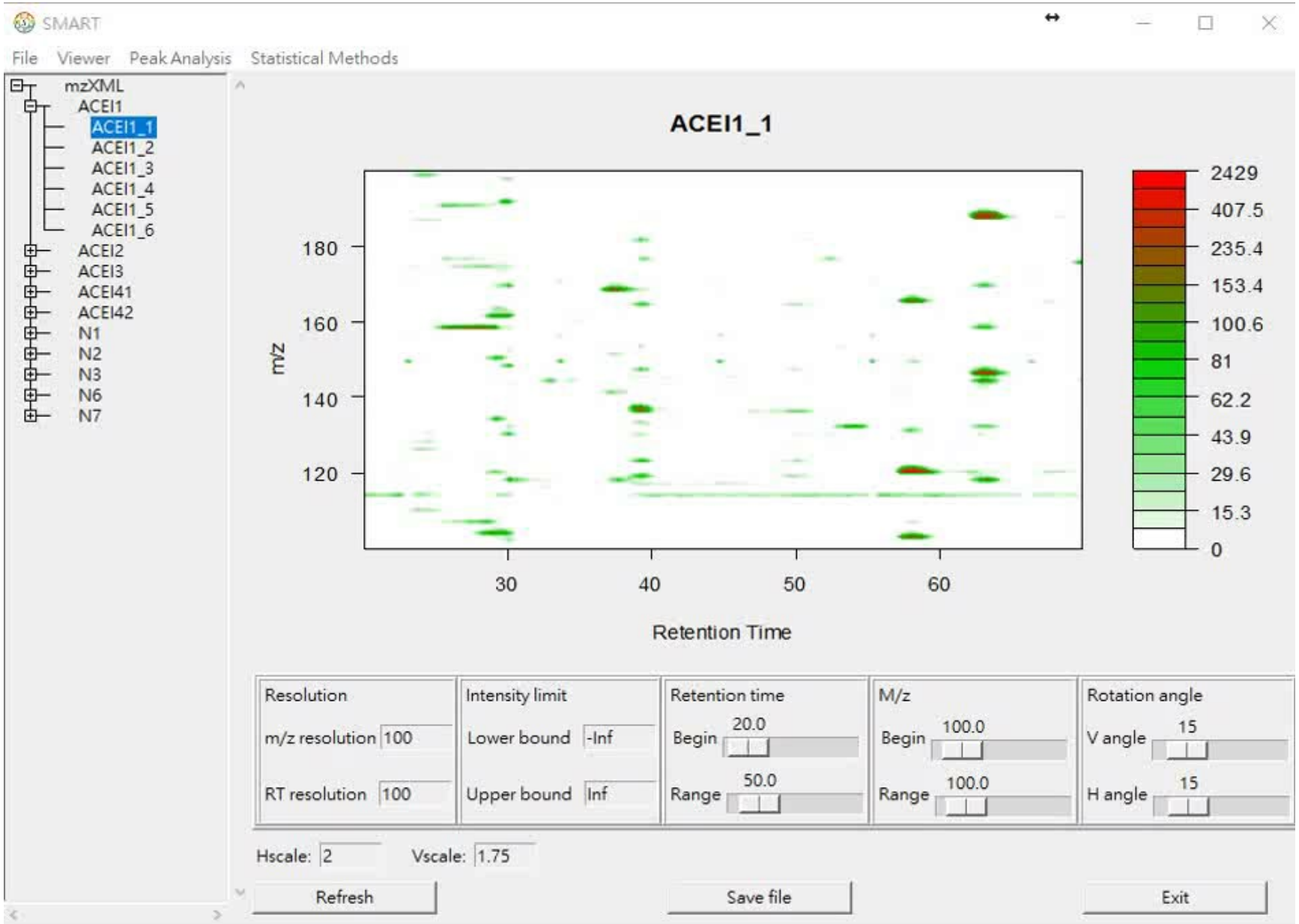


outpath\TIC plot\ACEI1_1_RT20(50).png

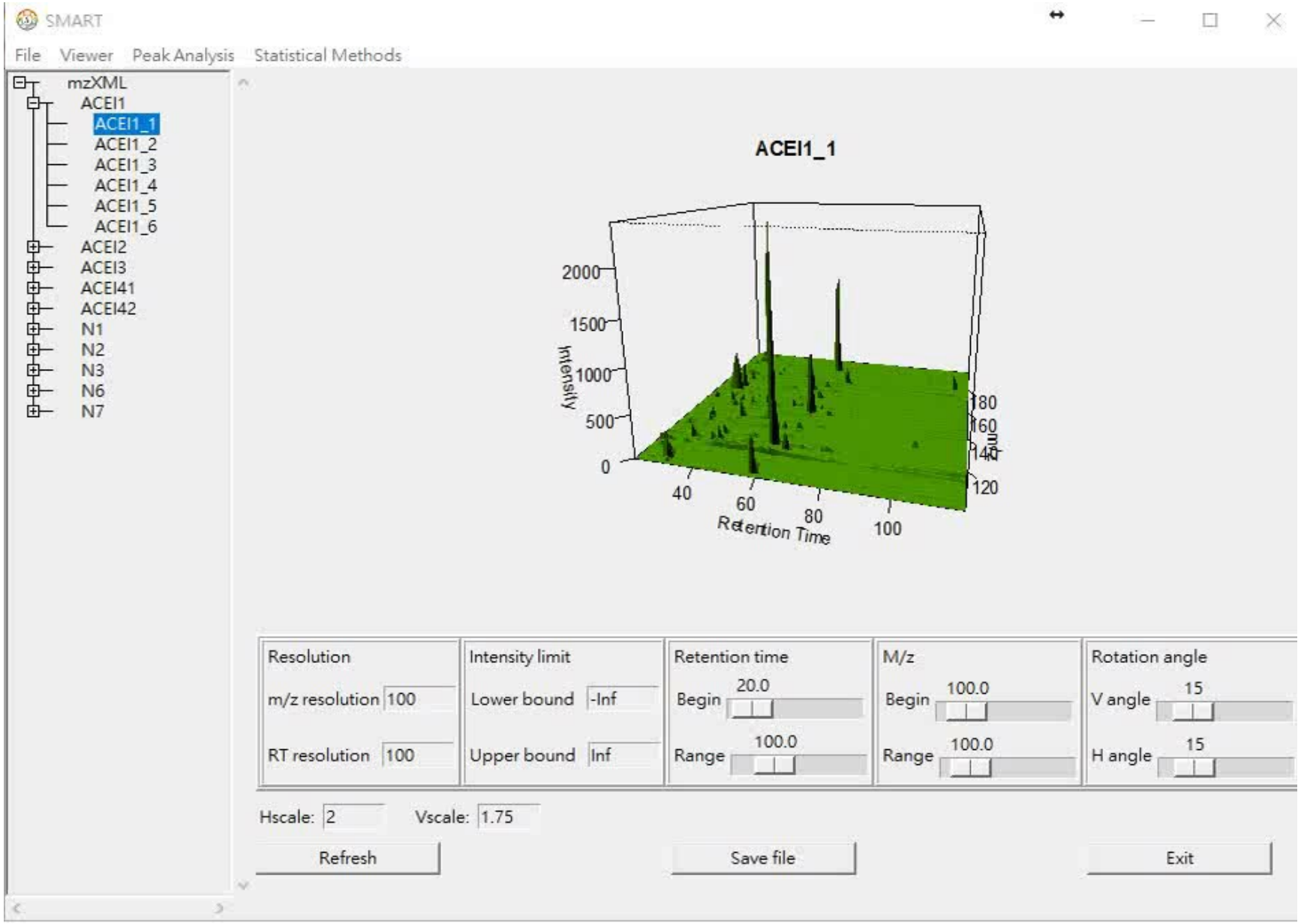



SMART
File
Viewer
Peak Analysis
Statistical Methods
mzXML
ACEI1
ACEI1_1
ACEI1_2
ACEI1_3
ACEI1_4
ACEI1_5
ACEI1_6
ACEI2
ACEI3
ACEI41
ACEI42
N1
N2
N3
N6
N7
Resolution
m/z resolution 100
RT resolution 100
Intensity limit
Lower bound -Inf
Upper bound Inf
Retention time
Begin 20
Range 50
M/z
Begin 100
Range 100
Rotation angle
V angle 15
H angle 15
Hscale: 2
Vscale: 1.75
Refresh
Save file
Exit

2D plot


SMART
File
Viewer
Peak Analysis
Statistical Methods
mzXML
ACEI1
ACEI1_1
ACEI1_2
ACEI1_3
ACEI1_4
ACEI1_5
ACEI1_6
ACEI2
ACEI3
ACEI41
ACEI42
N1
N2
N3
N6
N7
ACEI1_1
m/z
180
160
140
120
30
40
50
60
Retention Time
2429
407.5
235.4
153.4
100.6
81
62.2
43.9
29.6
15.3
0
Resolution
m/z resolution 100
RT resolution 100
Intensity limit
Lower bound -Inf
Upper bound Inf
Retention time
Begin 20.0
Range 50.0
M/z
Begin 100.0
Range 100.0
Rotation angle
V angle 15
H angle 15
Hscale: 2
Vscale: 1.75
Refresh
Save file
Exit

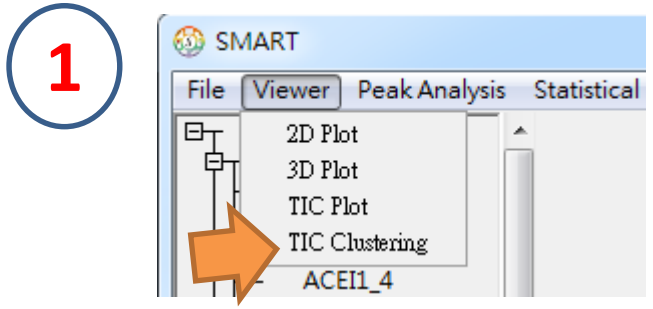
3D plot

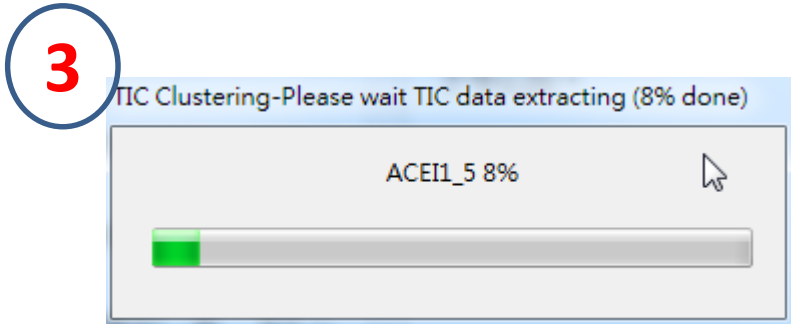

SMART
File
Viewer
Peak Analysis
Statistical Methods
mzXML
ACEI1
ACEI1_1
ACEI1_2
ACEI1_3
ACEI1_4
ACEI1_5
ACEI1_6
ACEI2
ACEI3
ACEI41
ACEI42
N1
N2
N3
N6
N7
ACEI1_1
Intensity
2000
1500
1000
500
0
40
60
80
100
Retention Time
180
160
140
120
Resolution
m/z resolution 100
RT resolution 100
Intensity limit
Lower bound -Inf
Upper bound Inf
Retention time
Begin 20.0
Range 100.0
M/z
Begin 100.0
Range 100.0
Rotation angle
V angle 15
H angle 15
Hscale: 2
Vscale: 1.75
Refresh
Save file
Exit

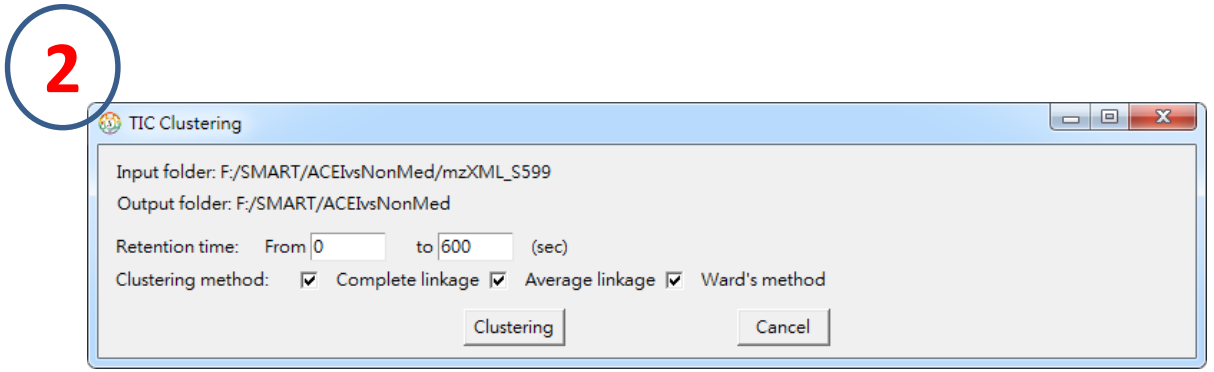
TIC plot

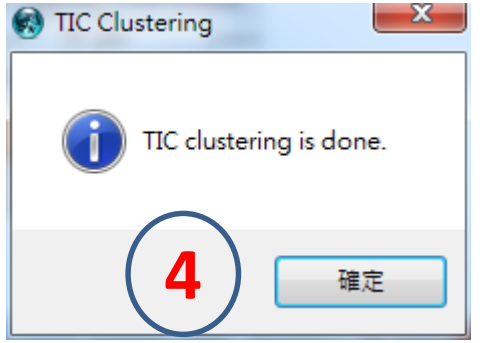
Data visualization

- TIC clustering









Statistical Metabolomics Analysis – an R Tool

- Data visualization
  - TIC clustering
    - output

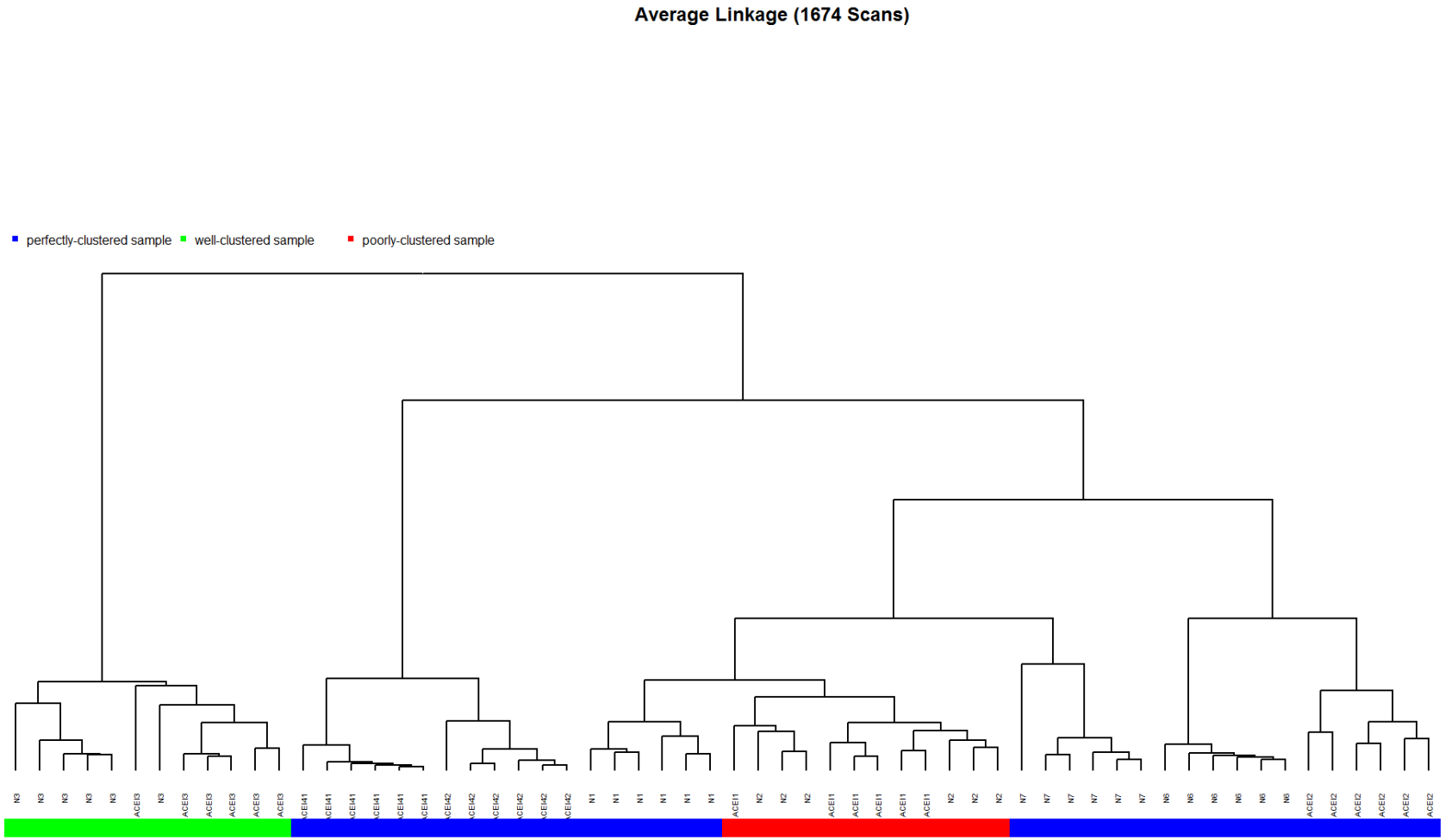
outpath\TIC clustering\TICtable.csv

	A	B	C	D	E	F	G	H	I	J	K
1	Scan_index	Ret_Time	ACEI1_1	ACEI1_2	ACEI1_3	ACEI1_4	ACEI1_5	ACEI1_6	ACEI2_1	ACEI2_2	ACEI2_3
2	1	1.611703	1840.316	1883.235	1999.769	1929.929	2029.317	2244.63	1590.959	1761.858	2098.871
3	2	1.914309	525.8319	513.814	509.7652	502.4335	512.8351	492.2585	480.4986	472.9005	519.8399
4	3	2.154968	562.8346	509.1748	539.4163	467.0927	511.4119	472.8939	543.9075	475.7264	537.953
5	4	2.395501	512.4183	508.6098	636.6686	503.0518	503.2116	464.3456	503.8277	548.2333	528.2615
6	5	2.635341	538.5349	520.6004	542.8813	545.037	507.3488	507.485	439.9774	494.179	526.823
7	6	2.875018	591.0371	545.6836	536.9702	542.5765	497.7138	488.2345	509.0443	535.7909	588.3315
8	7	3.114768	532.316	481.7373	572.3064	488.0685	531.7565	461.464	522.0878	490.4316	580.4105
9	8	3.35471	552.8936	549.1855	555.149	482.2098	464.2407	499.5836	528.0374	554.635	544.5468
10	9	3.594811	557.4039	485.2134	568.3623	560.4765	529.6015	451.1695	506.7829	518.2427	559.6974
11	10	3.83523	543.802	487.1033	569.2555	466.8808	459.995	447.294	493.1324	537.6599	556.3434
12	11	4.075452	540.6479	551.7796	583.8124	488.8864	514.7813	502.7545	502.8375	476.9177	565.6591
13	12	4.315626	520.1061	566.7769	545.4417	574.0717	494.1166	516.8445	468.7639	495.1702	515.7346
14	13	4.555596	565.7251	514.366	553.2607	503.1035	488.1889	516.7361	526.1669	523.666	564.338
15	14	4.795548	498.9484	464.8117	522.7676	527.2203	511.1384	473.0254	518.6355	529.5582	497.3565
16	15	5.035726	521.6926	494.5057	550.8154	514.8136	481.3621	501.2584	507.0366	539.1605	545.1652

Statistical Metabolomics Analysis – an R Tool

- Data visualization
  - TIC clustering
    - output

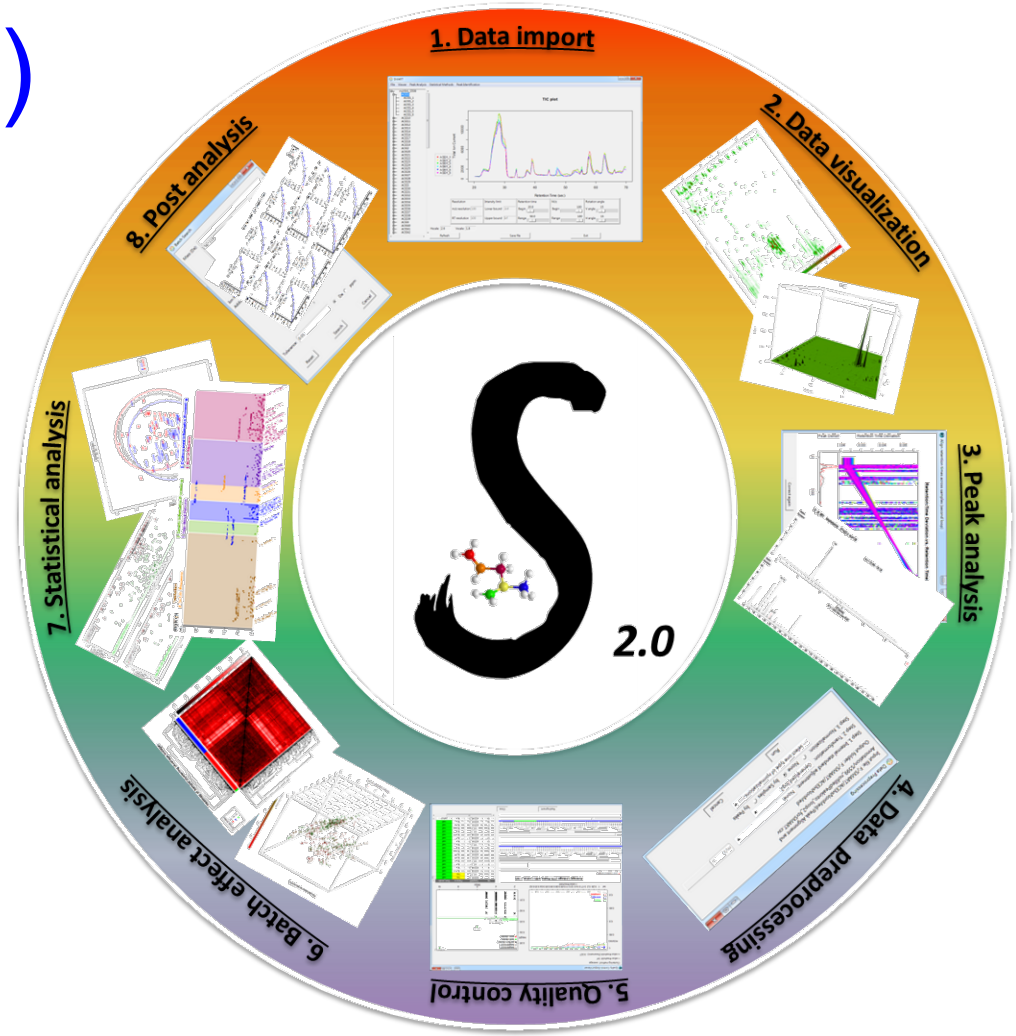
outpath\TIC clustering\Average\Clustering.png



TIC clustering

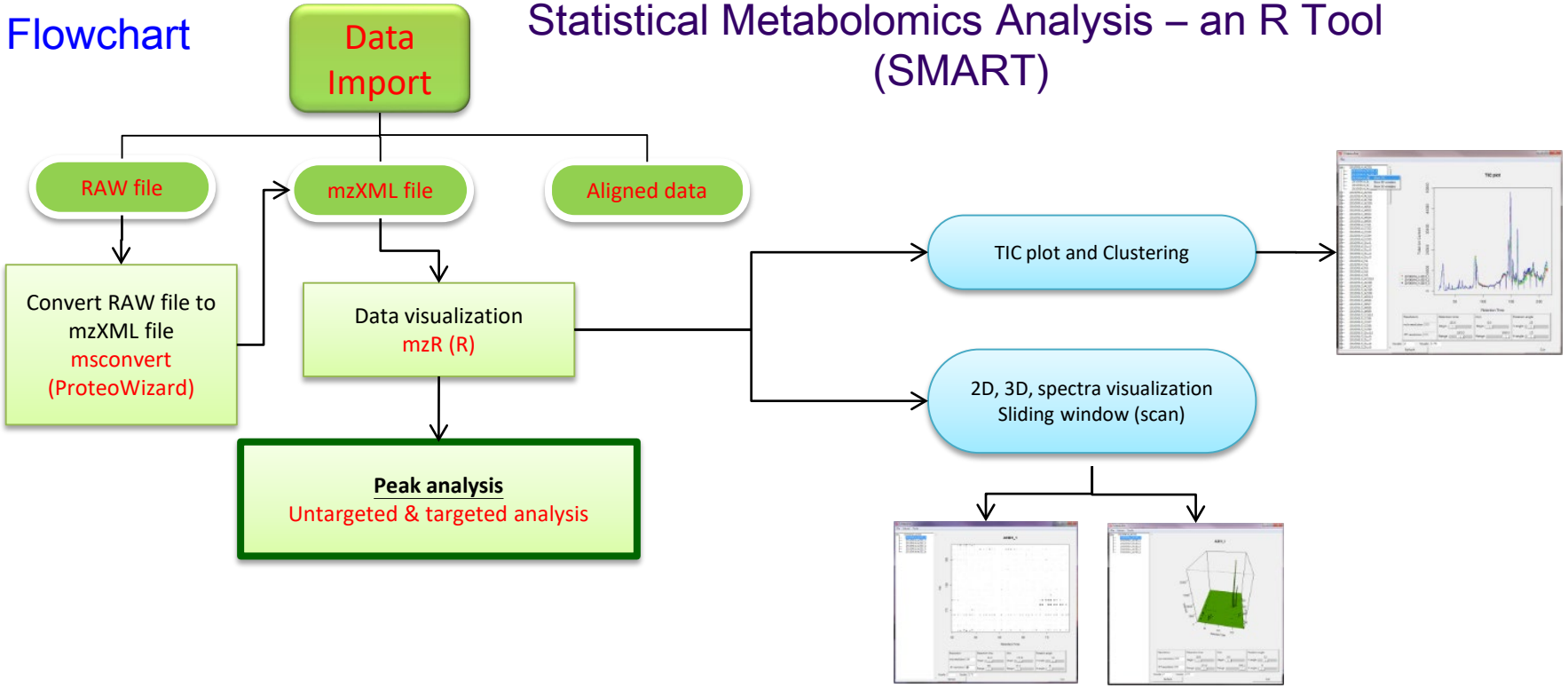
Statistical Metabolomics Analysis – an R Tool

3. Peak analysis (untargeted analysis)



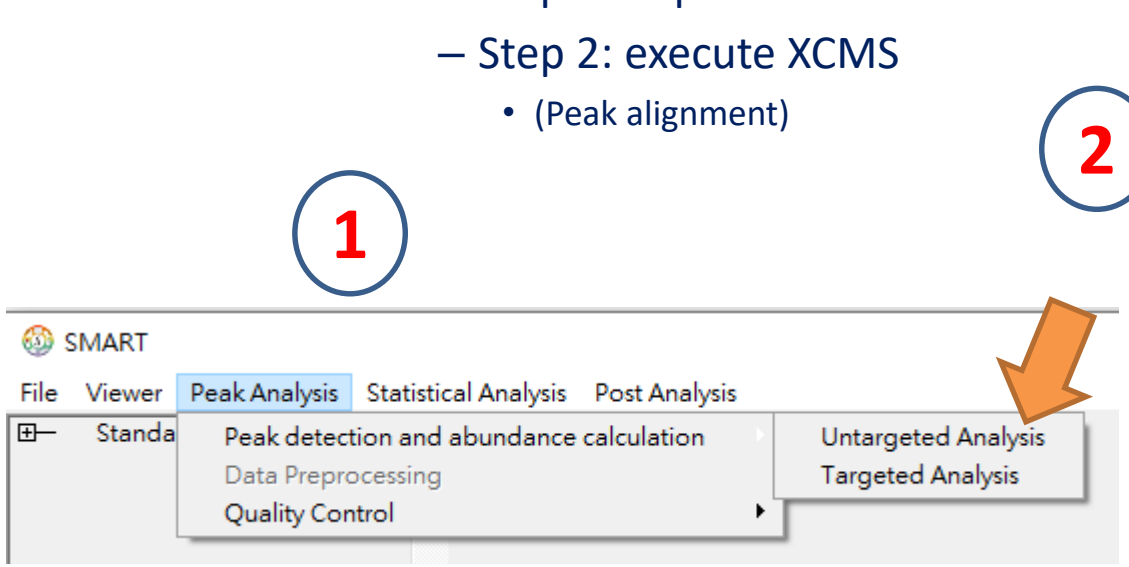
Statistical Metabolomics Analysis – an R Tool (SMART)

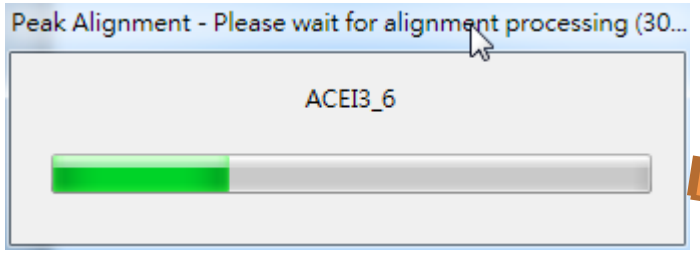
Flowchart

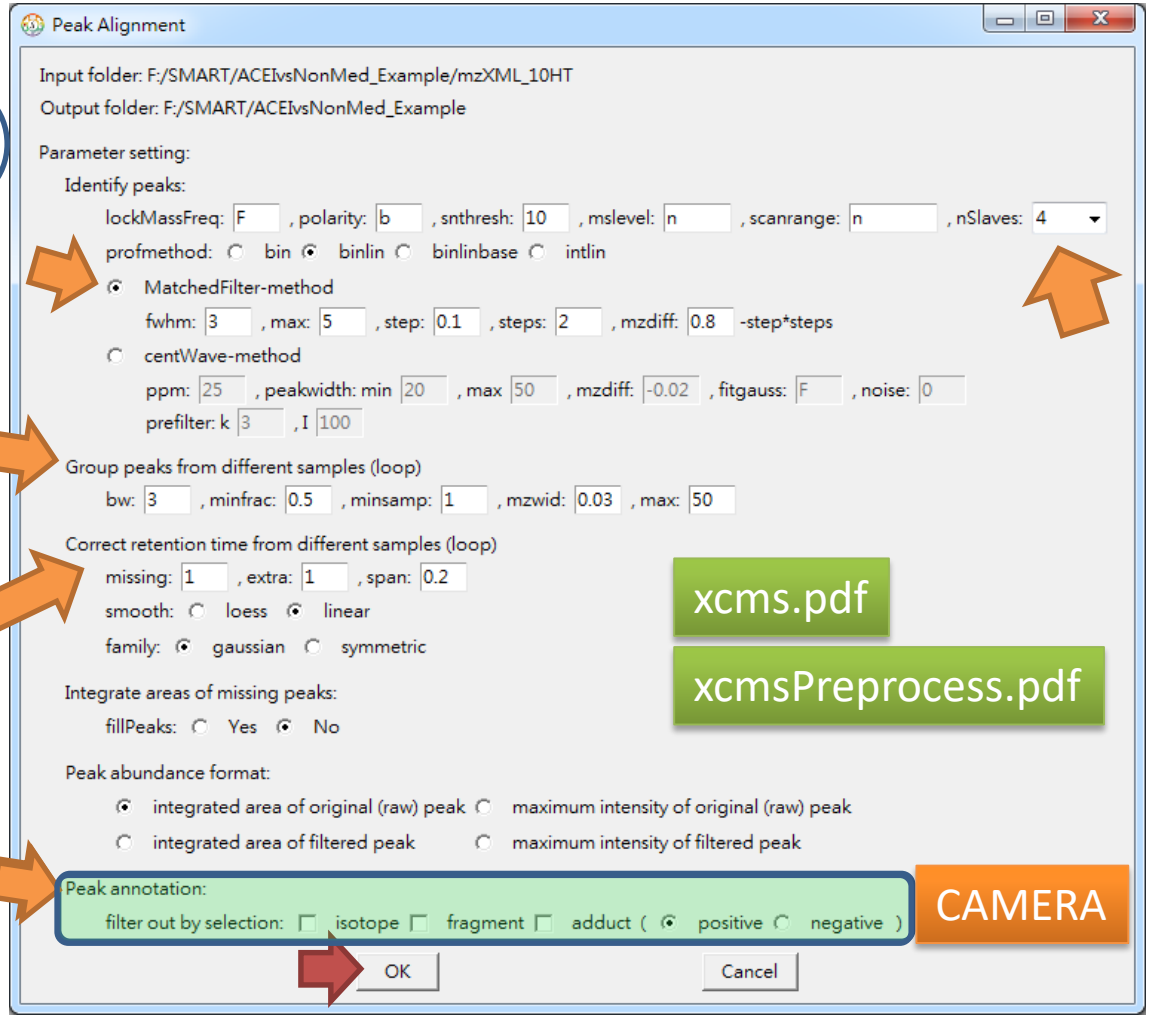


Peak analysis (untargeted)

- Untargeted analysis
  - Step 1: import mzXML files
  - Step 2: execute XCMS
    - (Peak alignment)

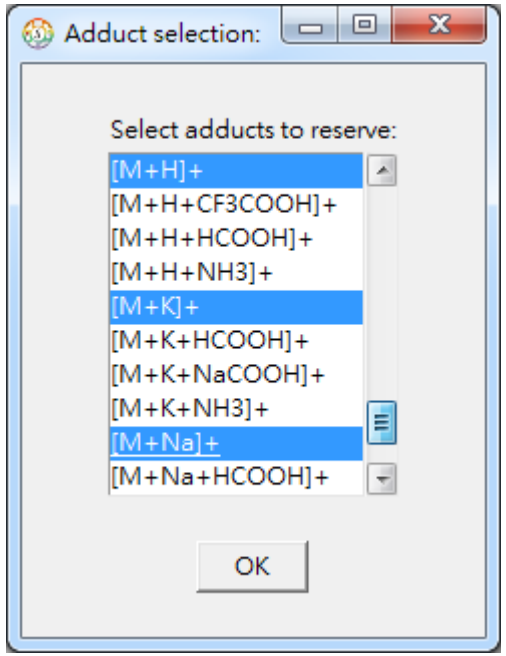


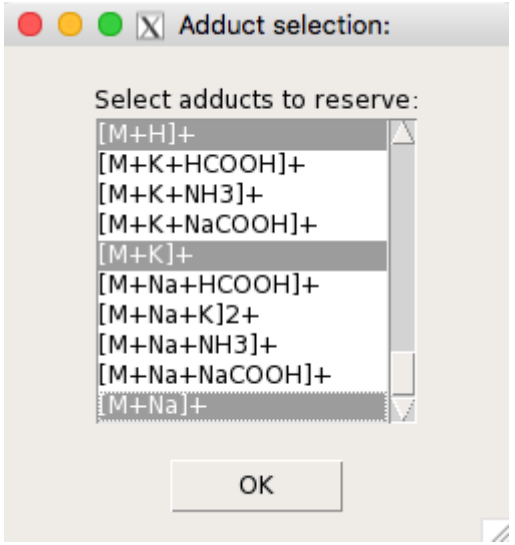




Peak analysis (untargeted)

– Step 3: execute CAMERA (Peak annotation)





Peak analysis (untargeted)

- 3. Peak alignment and annotation -- output file (peak abundance table)

outpath\Peak Alignment\S60_nofilledPeakinto_loop2.csv

Peak_Index	mz	Ret_Time.sec	ACEI1_1	ACEI1_2	ACEI1_3	ACEI1_4	ACEI1_5	ACEI1_6	ACEI2_1	ACEI2_2	ACEI2_3
1	55.9302	336.7816	40.49721	48.4055	68.5233	131.6603	153.4566	68.20429	98.66187	101.203	125.084
2	56.95207	335.9918	25.78232	NA	33.81898	31.96166	34.43379	36.10391	NA	21.12795	NA
3	57.93077	335.9468	25.30278	26.6211	29.52791	37.52189	34.32671	20.62405	19.76005	23.86996	29.22897
4	60.07351	29.2962	15.05838	19.10104	15.67263	20.25476	14.3002	19.61156	NA	17.66053	16.79027
5	68.97672	24.1171	32.99825	39.72613	41.26156	24.86147	37.08042	29.47383	48.53648	51.11766	60.06144
6	70.05911	32.3207	NA	NA	NA	NA	NA	NA	88.67861	71.20517	77.2781
7	72.07474	37.8148	77.36168	60.19769	53.10426	45.49223	52.31326	45.89031	266.037	248.2481	192.2031
8	72.07494	64.0368	NA	NA	NA	NA	NA	NA	61.4457	57.43003	54.12979
9	72.93235	336.7707	54.64176	NA	116.1767	105.3885	268.4261	NA	NA	116.3039	208.1015
10	74.93326	336.1583	NA	46.74469	53.22282	46.98308	44.75656	40.26063	50.46011	42.93352	57.50089
11	77.03265	57.9572	36.87367	38.9539	38.91873	31.10987	37.97594	34.42077	131.6498	148.9135	208.078
12	79.04987	57.9244	23.16134	15.64894	19.70598	19.23182	20.40361	15.09588	70.72101	60.65076	71.75067
13	82.00799	24.1165	22.04538	26.83705	28.74175	32.02328	22.70863	20.06136	41.06131	34.26637	40.51016
14	84.03872	30.0699	38.46723	45.41489	54.63141	50.46705	52.82711	49.82799	127.9401	147.1337	136.5419
15	84.03838	48.6612	33.55734	30.40641	21.84934	26.916	35.54772	27.8324	42.54551	47.40438	44.89901
16	84.95268	24.1224	245.3396	218.0391	247.8054	261.728	190.1302	217.5573	324.1442	256.0635	321.4737
17	84.95221	335.6901	301.2523	195.8762	316.6637	302.4194	272.9157	542.9502	232.0436	295.5475	193.4221
18	86.08994	53.8669	238.2553	219.8844	231.6355	185.6219	186.2405	190.8221	1209.354	1210.146	1024.845
19	86.98677	24.1066	17.84491	19.26533	NA	17.23194	20.42927	NA	33.63286	32.69955	25.17788
20	87.09331	53.7237	15.80857	NA	16.59535	NA	NA	NA	73.34223	76.04747	63.93756

Peak analysis (untargeted)

- 3. Peak alignment and annotation -- output file (peak abundance table)

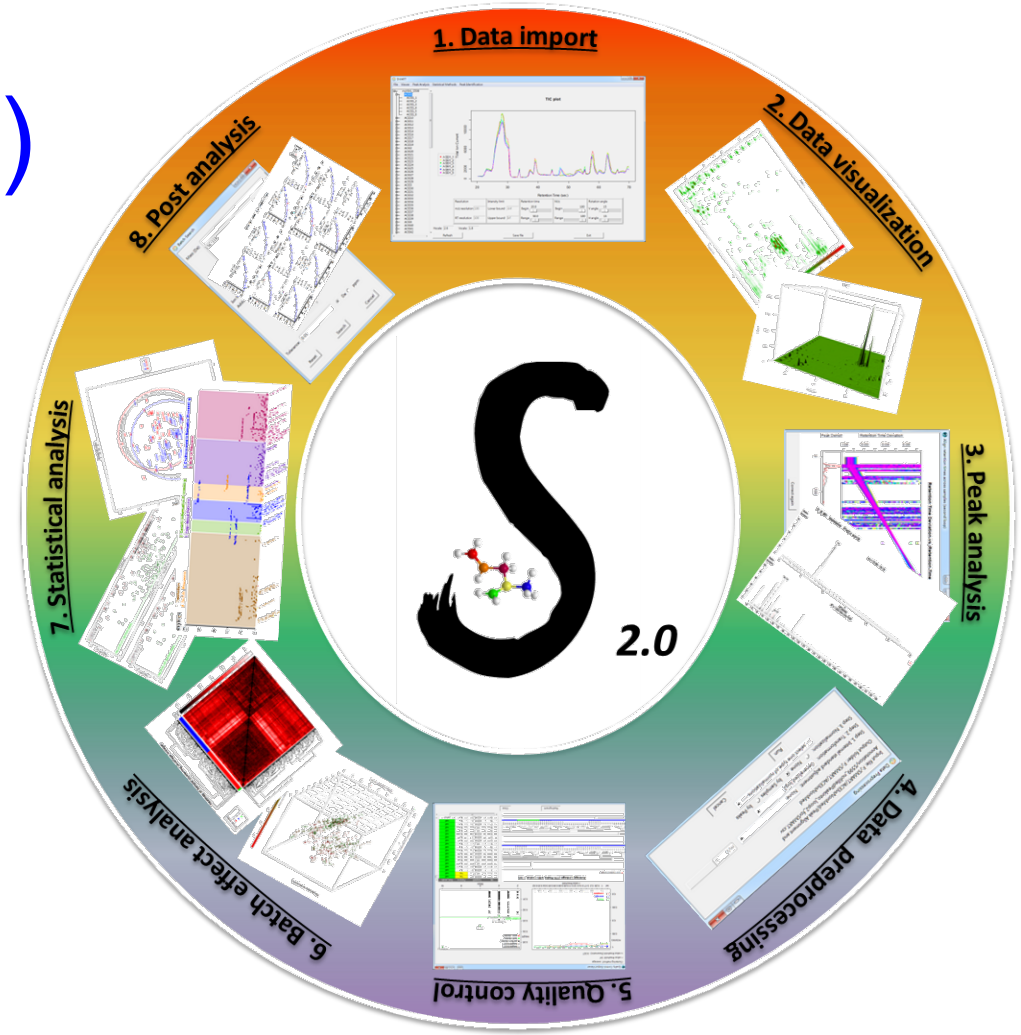
outpath\Peak Alignment\S60_nofilledPeakinto_loop2_annoinfo.csv

Peak_Index	mz	Ret_Time.sec	isotopes	adduct	pcgroup
1	22.9849305	27.7127			35
2	55.92994829	336.8763			319
3	56.95336493	335.936		[M+K]+ 17.9854	29
4	57.9299371	335.8817		[M+H]+ 56.9236	7
5	59.92473806	335.8649			220
6	60.07189387	29.25			336
7	68.97673761	24.1198		[M+Na]+ 45.9928	46
8	69.06404612	52.8241			182
9	70.05927741	32.3222			166
10	72.07471273	37.7522		[M+K]+ 33.1119 [M+Na]+ 49.0859 [M+H]+ 71.0677	44
11	72.07482812	64.0093			162
12	72.93202428	336.803		[M+Na]+ 49.9429	58
13	74.93307144	336.1642		[2M+K]+ 17.9854	29
14	76.92734109	336.1336		[3M+Na]+ 17.9854 [3M+Na+2K-H]2+ 17.9854	29
15	77.03232255	57.8967		[M+Na]+ 54.0431	4
16	79.04787429	57.9087			4
17	82.00795337	24.1116			46
18	84.03810375	30.1203			13
19	84.03824131	48.7586			252
20	84.07434463	38.9898		[M+H]+ 83.0689	23

Peak analysis (untargeted)

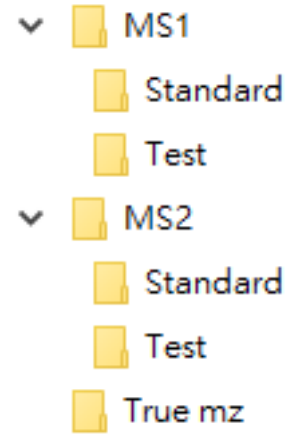
Statistical Metabolomics Analysis – an R Tool

3. Peak analysis (targeted analysis)

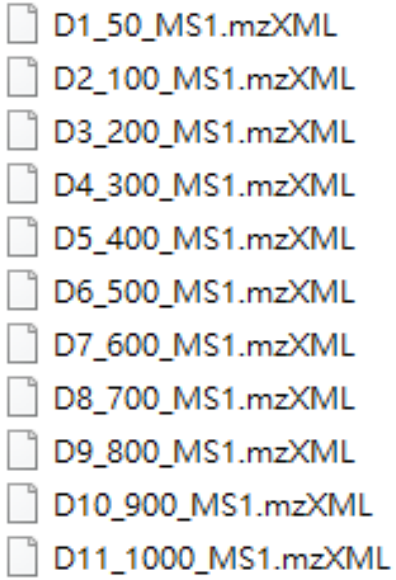


Peak analysis (targeted)

- Targeted analysis
  - Step 1: import mzXML files & true m/z information



1 mzXML files



2 true m/z

Name	mz	Ret_Time.sec	tol_time.sec
Heroin	370	685.2	5
Morphine	286	147	5
Cocaine	304	702	5
Thebaine	312	678.6	5
delta-9 THC	315	1513.8	5
Amphetamine	136	422.4	20
MA	150	495.6	10
MDMA	194	539.4	5
Love Drug	180	506.4	10
Ketamine	238	600	5
FM2	314	1044	5
Nimetazepam	296	1041	5

MS1

targeted_MS1_tol_time.csv

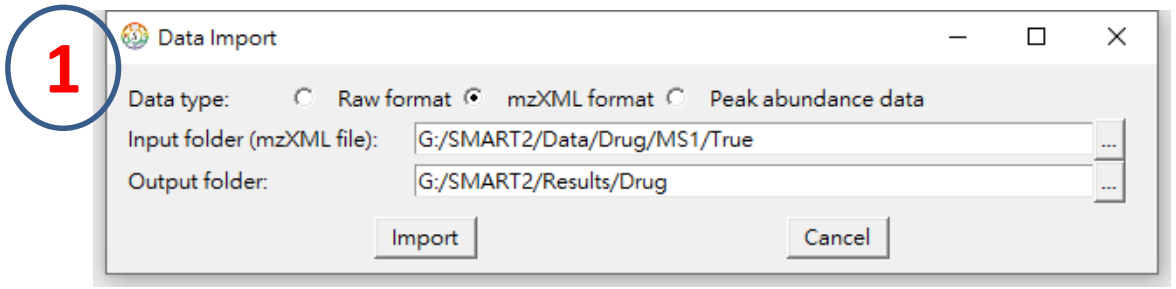
Name	parent_ion	daughter_ion	tol_time.sec
Heroin	370	370;328;268; 211;193;58	10
Morphine	286	286; 268; 229; 201	10
Cocaine	304	304; 182; 150; 105; 82	10
Thebaine	312	312; 281; 266; 251; 221; 58	10
delta-9 THC	315	315; 259; 193; 135; 123; 107; 93	10
Amphetamine	136	136; 119; 91	20
MA	150	150; 119; 91	10
MDMA	194	194; 163; 135; 133; 105	10
Love Drug	180	180; 163; 135; 133; 105	10
Ketamine	238	238; 220; 179; 163; 152; 125	10
FM2	314	314; 200; 286; 268	15
Nimetazepam	296	296; 268; 250; 222; 193; 165	10

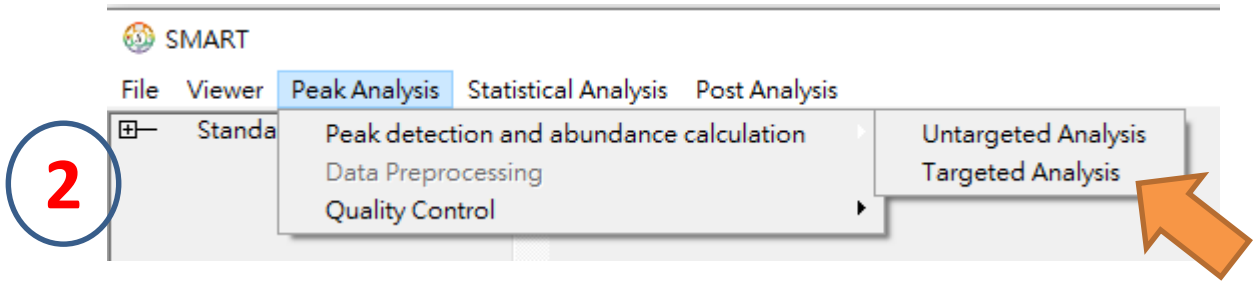
MS2

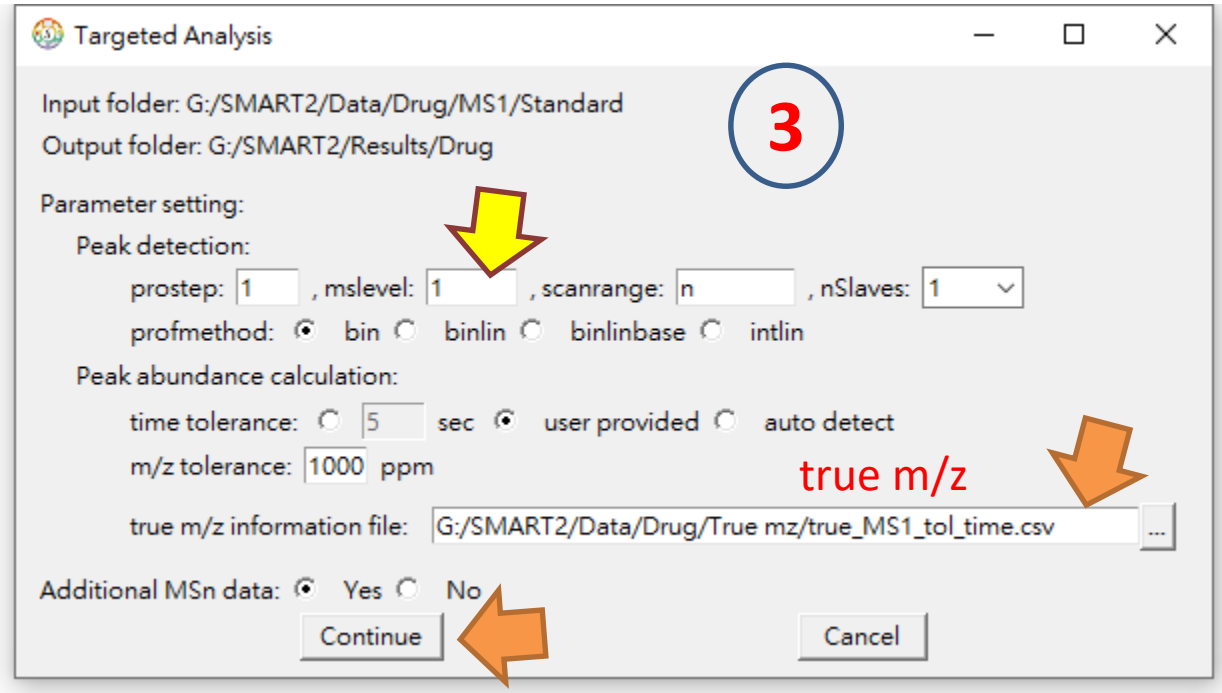
targeted_MS2_tol_time.csv

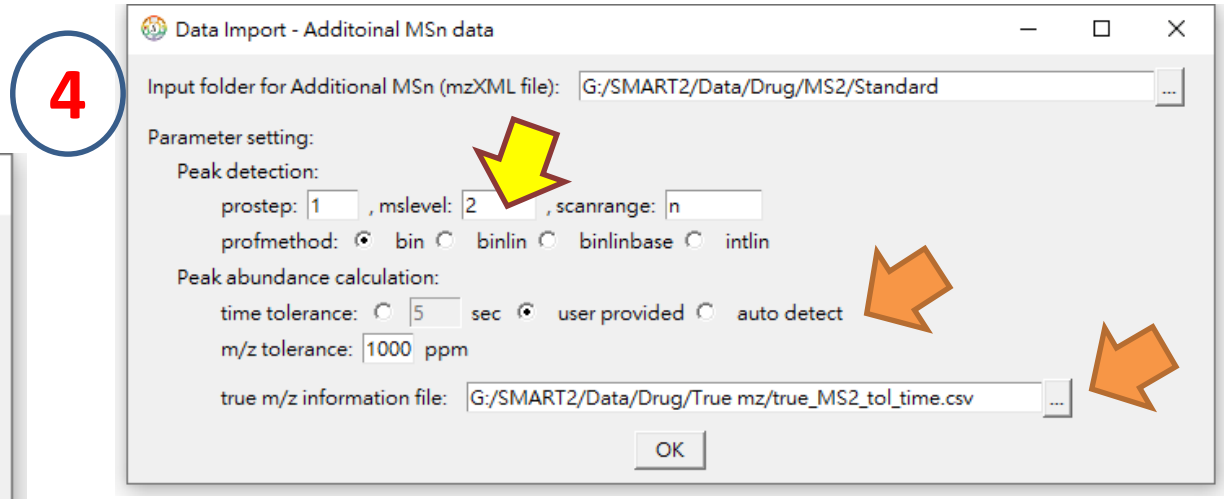
Peak analysis (targeted)

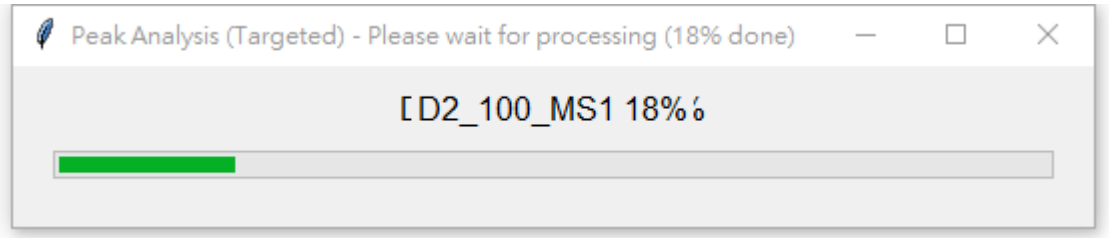
- Targeted analysis
  - Step 1: import mzXML files & correct m/z
  - Step 2: peak detection & abundance calculation







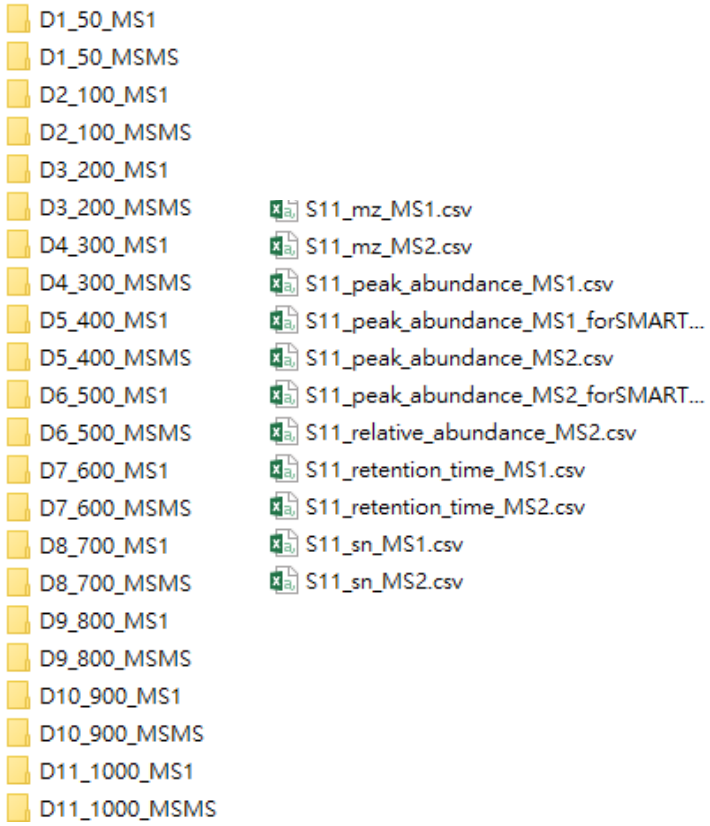


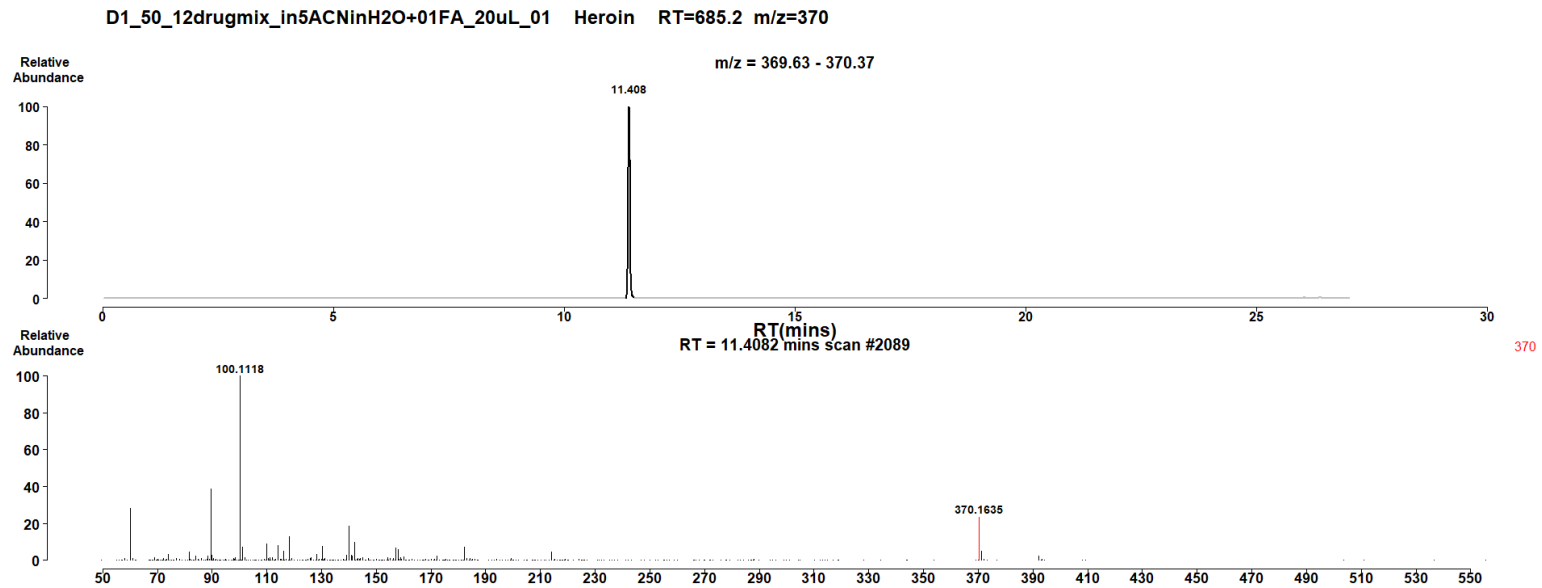


Peak analysis (targeted)

- Step 3. output file (peak abundance table; MS, MSMS spectrum)

outpath\Peak Analysis (Targeted)\S11_peak_abundance_MS.csv





Peak analysis (targeted)

Output file (Peak abundance table of all samples)

Peak abundance (MS1)

Peak_Index	Name	mz	Ret_Time.sec	Ret_Time_est.sec	mz_est	D1_50_MS1	D10_900_MS1	D11_1000_MS1	D2_100_MS1	D3_200_MS1	D4_300_MS1	D5_400_MS1	D6_500_MS1	D7_600_MS1	D8_700_MS1	D9_800_MS1
1	Heroin	370	685.2	685.3724	370.1643	20454233	4.85E+08	5.37E+08	33246830	65461545	1.08E+08	1.12E+08	2.2E+08	2.18E+08	3.07E+08	3.79E+08
2	Morphine	286	147	146.2696	286.1431	41215408	5.53E+08	6.23E+08	55632394	1.11E+08	1.59E+08	1.66E+08	2.97E+08	3.05E+08	4.01E+08	4.69E+08
3	Cocaine	304	702	701.136	304.1538	67391965	1.91E+09	2.16E+09	1.09E+08	2.8E+08	4.54E+08	4.79E+08	9.74E+08	9.95E+08	1.32E+09	1.57E+09
4	Thebaine	312	678.6	678.8176	312.1588	13918057	6.54E+08	7.9E+08	22156818	57057964	1.12E+08	1.14E+08	2.89E+08	3.04E+08	4.41E+08	5.71E+08
5	delta-9 THC	315	1513.8	1513.319	315.2302	584689	12324894	10750614	1040859	3017275	4419165	4255491	6385888	8052278	9572914	11385840
6	Amphetamine	136	422.4	419.1944	136.0804	28233349	38237498	37707935	30552141	36745807	37802678	38425018	38130787	39861533	37695091	38418264
7	MA	150	495.6	496.8365	150.1021	88366857	4.84E+08	5.19E+08	1.14E+08	1.84E+08	2.48E+08	2.57E+08	3.68E+08	3.78E+08	4.43E+08	4.61E+08
8	MDMA	194	539.4	538.9991	194.1173	57005641	1.36E+09	1.61E+09	77917415	1.73E+08	3.32E+08	3.11E+08	6.96E+08	7.22E+08	9.57E+08	1.14E+09
9	Love Drug	180	506.4	505.6407	180.0866	2977060	32927321	40623589	3508615	5215888	7552811	8060482	14862760	15982588	21387478	24410233
10	Ketamine	238	600	599.9313	238.0988	65361457	1.9E+09	2.12E+09	1.1E+08	3.04E+08	5.13E+08	5.38E+08	1.07E+09	1.13E+09	1.41E+09	1.64E+09
11	FM2	314	1044	1044.274	314.093	23235201	4.47E+08	5.37E+08	37148568	83573478	1.28E+08	1.31E+08	2.39E+08	2.43E+08	3.1E+08	3.84E+08
12	Nimetazepam	296	1041	1040.522	296.1024	27111684	5.03E+08	5.88E+08	41037216	91247724	1.44E+08	1.48E+08	2.57E+08	2.67E+08	3.45E+08	4.14E+08

Peak analysis (targeted)

Output file (Peak abundance table of all samples)

Peak abundance (MS2)

Peak_Index	Name	parent_ion	mz		Ret_Time.sec	Ret_Time_est.sec	mz_est	D1_50_MSMS	D10_900_MSMS	D11_1000_MSMS	D2_100_MSMS	D3_200_MSMS	D4_300_MSMS	D5_400_MSMS	D6_500_MSMS	D7_600_MSMS	D8_700_MSMS	D9_800_MSMS
1	Heroin	370	370	NA	685.5322	370.1636	34936037	4.31E+08	4.74E+08	45887062	80773701	1.19E+08	1.85E+08	2.31E+08	2.75E+08	2.99E+08	3.65E+08	
2	Heroin	370	328	NA	685.5322	328.1531	5994749	79250544	88352065	8205245	14903335	22352355	34701357	40951023	49833482	55034774	67134243	
3	Heroin	370	268	NA	685.5322	268.1322	5966182	80881111	96696971	8852504	15962231	23508212	36036774	44988809	54698078	57847453	70564375	
4	Heroin	370	211	NA	685.5322	211.0747	3361656	46230644	54063830	4770158	8791204	12473739	20439931	25000126	31336412	33862691	40365021	
5	Heroin	370	193	NA	685.5322	193.0641	1510510	20921156	23617575	2064699	3996665	5545420	8870916	9727212	12878930	14027716	17201483	
6	Heroin	370	58	NA	686.438	57.99157	NA	258.2116	NA	273.661	NA	NA	638.7396	NA	NA	NA	NA	
7	Morphine	286	286	NA	145.5	286.143	16131323	2.1E+08	2.22E+08	34800087	63534132	84246919	1.1E+08	1.36E+08	1.46E+08	1.61E+08	1.97E+08	
8	Morphine	286	268	NA	145.5	268.1305	430782.9	6575950	6785043	969792.4	1816443	2392807	3283501	4052850	4471966	4903943	6066029	
9	Morphine	286	229	NA	145.5	229.0854	1381069	19277741	20679948	3056907	5480279	7688644	10079360	11702549	13521425	14964256	17815511	
10	Morphine	286	201	NA	145.5	201.0879	1967459	28574132	28953802	4338963	8250409	11269557	14353928	17231385	19143587	21127442	25307275	
11	Cocaine	304	304	NA	700.1591	304.1536	1869106	18330609	20144319	4651245	7718511	9179664	12901861	13487416	15494409	16525349	17790183	
12	Cocaine	304	182	NA	699.8602	182.117	43752791	5.02E+08	5.37E+08	1.18E+08	1.99E+08	2.42E+08	3.23E+08	3.65E+08	4.08E+08	4.44E+08	4.85E+08	
13	Cocaine	304	150	NA	699.8602	150.091	4216007	54187027	58594892	12328872	21144421	26080012	33884162	37684106	42701985	46630588	51139926	
14	Cocaine	304	105	NA	699.8602	105.0332	3613007	50772687	56131246	10074624	17121746	21047970	29647120	34590745	35860013	40174301	46894619	
15	Cocaine	304	82	NA	699.8602	82.0649	5027314	69164979	78411520	14017254	23880243	29053601	40693701	46962217	51254207	56796007	64853128	
16	Thebaine	312	312	NA	678.0616	312.1517	1095431	9256829	11341677	1766889	3150661	3824506	5264614	6189062	7429242	7942980	8623290	
17	Thebaine	312	281	NA	678.0616	281.1163	534162.5	4365164	4921961	804912.5	1398028	1823320	2546612	2982317	3328803	3572010	3994250	
18	Thebaine	312	266	NA	678.0616	266.1032	1866305	16921521	19156966	3263283	5384742	6937593	9418883	11512846	12833181	13809395	15751537	
19	Thebaine	312	251	NA	678.0616	251.0695	3056593	26949412	31532800	5253820	8837063	11769548	15247149	18110674	21036849	23342167	25431602	
20	Thebaine	312	221	NA	678.0616	221.0838	1812587	17278121	19202136	3174271	5577776	7159779	9476728	11502196	12895684	14336249	15762883	
21	Thebaine	312	58	NA	678.7908	58.04143	NA	5888.264	7090.007	4134.432	286.0357	4603.693	14751.52	6566.844	12127.54	11897.7	7630.39	

Peak analysis (targeted)

Output file

Relative abundance (MS2)

Peak_Index	Name	parent_ion	mz	Ret_Time.sec	Ret_Time_est.sec	mz_est	D1_50_MSMS	D10_900_MSMS	D11_1000_MSMS	D2_100_MSMS	D3_200_MSMS	D4_300_MSMS	D5_400_MSMS	D6_500_MSMS	D7_600_MSMS	D8_700_MSMS	D9_800_MSMS
1	Heroin	370	370	NA	685.5322	370.1636	100	100	100	100	100	100	100	100	100	100	100
2	Heroin	370	328	NA	685.5322	328.1531	17.15921	18.39443	18.6417	17.88139	18.45073	18.73004	18.74169	17.76055	18.09249	18.37667	18.39403
3	Heroin	370	268	NA	685.5322	268.1322	17.07744	18.77289	20.40242	19.29194	19.76167	19.69859	19.46293	19.51175	19.85863	19.31585	19.33385
4	Heroin	370	211	NA	685.5322	211.0747	9.622315	10.73035	11.40711	10.39543	10.88375	10.45231	11.0393	10.84261	11.37696	11.3071	11.05956
5	Heroin	370	193	NA	685.5322	193.0641	4.323645	4.855901	4.983152	4.499524	4.947978	4.646757	4.79105	4.218714	4.675811	4.683997	4.713014
6	Heroin	370	58	NA	686.438	57.99157	NA	5.99E-05	NA	0.000596	NA	NA	0.000345	NA	NA	NA	NA
7	Morphine	286	286	NA	145.5	286.143	100	100	100	100	100	100	100	100	100	100	100
8	Morphine	286	268	NA	145.5	268.1305	2.670475	3.125105	3.062768	2.786753	2.859004	2.840231	2.987035	2.972688	3.06945	3.042394	3.075344
9	Morphine	286	229	NA	145.5	229.0854	8.561415	9.161407	9.334928	8.784194	8.625725	9.126321	9.169298	8.583596	9.280783	9.283786	9.032074
10	Morphine	286	201	NA	145.5	201.0879	12.19651	13.57935	13.06975	12.46825	12.98579	13.37682	13.05792	12.63889	13.1397	13.10741	12.83023
11	Cocaine	304	304	NA	700.1591	304.1536	4.271969	3.654745	3.751795	3.939854	3.871665	3.786899	3.996785	3.692513	3.800112	3.720228	3.665711
12	Cocaine	304	182	NA	699.8602	182.117	100	100	100	100	100	100	100	100	100	100	100
13	Cocaine	304	150	NA	699.8602	150.091	9.635973	10.80377	10.91305	10.44322	10.60621	10.75882	10.49676	10.31695	10.47296	10.4976	10.53751
14	Cocaine	304	105	NA	699.8602	105.0332	8.257776	10.12303	10.45421	8.533747	8.588402	8.682946	9.184191	9.47007	8.794919	9.04414	9.662752
15	Cocaine	304	82	NA	699.8602	82.0649	11.49027	13.79007	14.60382	11.87337	11.97852	11.98552	12.60624	12.85707	12.57045	12.78606	13.36315

Peak analysis (targeted)

Output file

m/z (MS1)

Peak_Index	Name	mz	Ret_Time.sec	Ret_Time_est.sec	mz_est	D1_50_MS1	D10_900_MS1	D11_1000_MS1	D2_100_MS1	D3_200_MS1	D4_300_MS1	D5_400_MS1	D6_500_MS1	D7_600_MS1	D8_700_MS1	D9_800_MS1
1	Heroin	370	685.2	685.3724	370.1643	370.1641	370.1642	370.1642	370.1643	370.1644	370.1643	370.1643	370.1644	370.1643	370.1642	370.1643
2	Morphine	286	147	146.2696	286.1431	286.1432	286.1431	286.143	286.1432	286.1432	286.1432	286.1432	286.1432	286.1431	286.1431	286.1431
3	Cocaine	304	702	701.136	304.1538	304.1537	304.1538	304.1538	304.1539	304.154	304.1538	304.1537	304.1539	304.1538	304.1539	304.1538
4	Thebaine	312	678.6	678.8176	312.1588	312.1588	312.1587	312.1586	312.1588	312.1588	312.1588	312.1588	312.1588	312.1588	312.1587	312.1587
5	delta-9 THC	315	1513.8	1513.319	315.2302	315.2193	315.2314	315.2317	315.2313	315.2314	315.2313	315.2312	315.2313	315.2311	315.2312	315.2313
6	Amphetamine	136	422.4	419.1944	136.0804	136.0589	136.0881	136.0891	136.0615	136.0743	136.0787	136.0822	136.0876	136.0871	136.088	136.0886
7	MA	150	495.6	496.8365	150.1021	150.0957	150.1096	150.1148	150.0878	150.0931	150.0988	150.0962	150.1067	150.0988	150.1092	150.1121
8	MDMA	194	539.4	538.9991	194.1173	194.1173	194.1172	194.1172	194.1171	194.1173	194.1173	194.1173	194.1173	194.1173	194.1173	194.1172
9	Love Drug	180	506.4	505.6407	180.0866	180.0712	180.0972	180.101	180.0718	180.084	180.0921	180.0782	180.0869	180.0861	180.0898	180.094
10	Ketamine	238	600	599.9313	238.0988	238.0988	238.099	238.0989	238.0988	238.0988	238.0988	238.0988	238.0988	238.0988	238.0988	238.099
11	FM2	314	1044	1044.274	314.093	314.0929	314.093	314.093	314.0932	314.0931	314.093	314.0931	314.0929	314.093	314.093	314.093
12	Nimetazepam	296	1041	1040.522	296.1024	296.1024	296.1025	296.1025	296.1024	296.1024	296.1024	296.1025	296.1022	296.1024	296.1023	296.1026

Peak analysis (targeted)

Output file

m/z (MS2)

Peak_Index	Name	parent_ion	mz	Ret_Time.sec	Ret_Time_est.sec	mz_est	D1_50_MSMS	D10_900_MSMS	D11_1000_MSMS	D2_100_MSMS	D3_200_MSMS	D4_300_MSMS	D5_400_MSMS	D6_500_MSMS	D7_600_MSMS	D8_700_MSMS	D9_800_MSMS
1	Heroin	370	370	NA	685.5322	370.1636	370.1635	370.1637	370.1635	370.1635	370.1635	370.1634	370.1635	370.1636	370.1638	370.1635	370.1636
2	Heroin	370	328	NA	685.5322	328.1531	328.1533	328.1531	328.1528	328.1533	328.1532	328.1531	328.1531	328.153	328.1531	328.1531	328.153
3	Heroin	370	268	NA	685.5322	268.1322	268.1324	268.1322	268.132	268.1324	268.1323	268.1323	268.1322	268.1322	268.1322	268.1321	268.1322
4	Heroin	370	211	NA	685.5322	211.0747	211.0746	211.0746	211.0746	211.0745	211.0747	211.0748	211.0748	211.0748	211.0746	211.0746	211.0746
5	Heroin	370	193	NA	685.5322	193.0641	193.0642	193.0642	193.064	193.0641	193.0642	193.0641	193.0641	193.0641	193.0641	193.0641	193.0642
6	Heroin	370	58	NA	686.438	57.99157	NA	57.9707	NA	57.9983	NA	NA	58.0057	NA	NA	NA	NA
7	Morphine	286	286	NA	145.5	286.143	286.1429	286.1428	286.1428	286.1431	286.143	286.1431	286.1431	286.143	286.1431	286.1429	286.1429
8	Morphine	286	268	NA	145.5	268.1305	268.1101	268.1326	268.1325	268.1327	268.1326	268.1326	268.1326	268.1325	268.1326	268.1325	268.1325
9	Morphine	286	229	NA	145.5	229.0854	229.0853	229.0854	229.0854	229.0854	229.0854	229.0854	229.0854	229.0854	229.0855	229.0854	229.0854
10	Morphine	286	201	NA	145.5	201.0879	201.0616	201.0905	201.0905	201.0905	201.0905	201.0904	201.0905	201.0905	201.0906	201.0906	201.0904
11	Cocaine	304	304	NA	700.1591	304.1536	304.1536	304.1536	304.1536	304.1537	304.1537	304.1534	304.1538	304.1538	304.1537	304.1537	304.1535
12	Cocaine	304	182	NA	699.8602	182.117	182.1171	182.117	182.117	182.117	182.117	182.1169	182.117	182.117	182.1171	182.117	182.1171
13	Cocaine	304	150	NA	699.8602	150.091	150.0909	150.091	150.0909	150.091	150.091	150.0909	150.091	150.091	150.091	150.0909	150.0909
14	Cocaine	304	105	NA	699.8602	105.0332	105.0332	105.0332	105.0332	105.0332	105.0332	105.0332	105.0333	105.0332	105.0332	105.0332	105.0332
15	Cocaine	304	82	NA	699.8602	82.0649	82.0649	82.0649	82.0649	82.0649	82.0649	82.0649	82.0649	82.0649	82.0649	82.0649	82.0649

Peak analysis (targeted)

Output file

RT (MS1)

Peak_Index	Name	mz	Ret_Time.sec	Ret_Time_est.sec	mz_est	D1_50_MS1	D10_900_MS1	D11_1000_MS1	D2_100_MS1	D3_200_MS1	D4_300_MS1	D5_400_MS1	D6_500_MS1	D7_600_MS1	D8_700_MS1	D9_800_MS1
1	Heroin	370	685.2	685.3724	370.1643	684.492	685.62	685.626	685.392	684.96	685.428	685.062	685.68	685.632	685.89	685.314
2	Morphine	286	147	146.2696	286.1431	144.954	146.202	147.018	145.902	146.196	146.196	146.202	146.154	146.796	146.85	146.496
3	Cocaine	304	702	701.136	304.1538	699.912	701.256	701.226	700.8	701.274	701.034	701.286	701.43	701.394	701.586	701.298
4	Thebaine	312	678.6	678.8176	312.1588	678.264	678.828	678.858	678.504	678.786	678.672	678.948	678.744	679.002	679.314	679.074
5	delta-9 THC	315	1513.8	1513.319	315.2302	1513.5	1513.308	1513.782	1512.63	1513.332	1512.768	1513.47	1513.41	1513.632	1513.422	1513.26
6	Amphetamine	136	422.4	419.1944	136.094	421.452	415.902	414.804	423.384	420.396	418.092	417.762	416.316	430.752	416.67	415.608
7	MA	150	495.6	496.8365	150.1021	495.9	496.668	496.824	496.854	496.818	496.458	496.794	495.954	498.144	497.148	497.64
8	MDMA	194	539.4	538.9991	194.1173	538.212	539.214	539.298	538.506	538.752	538.824	539.424	539.172	539.496	538.992	539.1
9	Love Drug	180	506.4	505.6407	180.0866	504.756	505.356	506.364	504.726	505.014	505.836	506.154	505.44	506.694	505.932	505.776
10	Ketamine	238	600	599.9313	238.0988	599.22	599.994	600.06	599.826	599.922	599.88	600.156	599.796	600.072	600.12	600.198
11	FM2	314	1044	1044.274	314.093	1044.162	1044.828	1044.162	1043.718	1044.252	1044.438	1043.772	1044.33	1044.582	1044.99	1043.778
12	Nimetazepam	296	1041	1040.522	296.1024	1039.692	1041.312	1040.952	1040.52	1040.508	1040.46	1040.1	1040.442	1040.688	1040.538	1040.532

Peak analysis (targeted)

Output file

RT (MS2)

Peak_Index	Name	parent_ion	mz	Ret_Time.sec	Ret_Time_est.sec	mz_est	D1_50_MSMS	D10_900_MSMS	D11_1000_MSMS	D2_100_MSMS	D3_200_MSMS	D4_300_MSMS	D5_400_MSMS	D6_500_MSMS	D7_600_MSMS	D8_700_MSMS	D9_800_MSMS
1	Heroin	370	370	NA	685.5322	370.1636	683.292	686.46	686.472	683.16	683.1	686.37	686.376	686.4	686.376	686.424	686.424
2	Heroin	370	328	NA	685.5322	328.1531	683.292	686.46	686.472	683.16	683.1	686.37	686.376	686.4	686.376	686.424	686.424
3	Heroin	370	268	NA	685.5322	268.1322	683.292	686.46	686.472	683.16	683.1	686.37	686.376	686.4	686.376	686.424	686.424
4	Heroin	370	211	NA	685.5322	211.0747	683.292	686.46	686.472	683.16	683.1	686.37	686.376	686.4	686.376	686.424	686.424
5	Heroin	370	193	NA	685.5322	193.0641	683.292	686.46	686.472	683.16	683.1	686.37	686.376	686.4	686.376	686.424	686.424
6	Heroin	370	58	NA	686.438	57.99157	NA	679.944	NA	689.724	NA	NA	689.646	NA	NA	NA	NA
7	Morphine	286	286	NA	145.5	286.143	145.524	145.518	145.494	145.47	145.5	145.488	145.482	145.5	145.5	145.524	145.5
8	Morphine	286	268	NA	145.5	268.1305	145.524	145.518	145.494	145.47	145.5	145.488	145.482	145.5	145.5	145.524	145.5
9	Morphine	286	229	NA	145.5	229.0854	145.524	145.518	145.494	145.47	145.5	145.488	145.482	145.5	145.5	145.524	145.5
10	Morphine	286	201	NA	145.5	201.0879	145.524	145.518	145.494	145.47	145.5	145.488	145.482	145.5	145.5	145.524	145.5
11	Cocaine	304	304	NA	700.1591	304.1536	700.038	699.876	703.164	699.906	699.816	699.816	699.828	699.84	699.792	699.84	699.834
12	Cocaine	304	182	NA	699.8602	182.117	700.038	699.876	699.876	699.906	699.816	699.816	699.828	699.84	699.792	699.84	699.834
13	Cocaine	304	150	NA	699.8602	150.091	700.038	699.876	699.876	699.906	699.816	699.816	699.828	699.84	699.792	699.84	699.834
14	Cocaine	304	105	NA	699.8602	105.0332	700.038	699.876	699.876	699.906	699.816	699.816	699.828	699.84	699.792	699.84	699.834
15	Cocaine	304	82	NA	699.8602	82.0649	700.038	699.876	699.876	699.906	699.816	699.816	699.828	699.84	699.792	699.84	699.834

Peak analysis (targeted)

Output file

S/N (MS1)

Peak_Index	Name	mz	Ret_Time.sec	Ret_Time_est.sec	mz_est	D1_50_MS1	D10_900_MS1	D11_1000_MS1	D2_100_MS1	D3_200_MS1	D4_300_MS1	D5_400_MS1	D6_500_MS1	D7_600_MS1	D8_700_MS1	D9_800_MS1
1	Heroin	370	685.2	685.3724	370.1643	1.258	1.3451	1.3331	1.2637	1.2006	1.2379	1.2124	1.3227	1.2697	1.2882	1.3332
2	Morphine	286	147	146.2696	286.1431	0.8601	0.8815	0.8624	0.7993	0.7657	0.7593	0.7614	0.8325	0.821	0.889	0.8425
3	Cocaine	304	702	701.136	304.1538	1.0586	1.1744	1.1824	1.0686	1.1593	1.1368	1.1814	1.1737	1.1954	1.1696	1.1643
4	Thebaine	312	678.6	678.8176	312.1588	1.2042	1.3537	1.3344	1.1617	1.2331	1.3413	1.3679	1.4252	1.4004	1.35	1.39
5	delta-9 THC	315	1513.8	1513.319	315.2302	0.4827	0.7615	0.786	0.6074	0.7792	0.7682	0.7972	0.812	0.8056	0.7853	0.8336
6	Amphetamine	136	422.4	419.1944	136.0804	1.4295	0.8961	0.8624	1.4094	1.2505	1.1382	1.1193	0.9806	0.9936	0.9322	0.9072
7	MA	150	495.6	496.8365	150.1021	1.2172	1.0342	1.0247	1.1924	1.1465	1.1408	1.1538	1.1191	1.1219	1.0873	1.0896
8	MDMA	194	539.4	538.9991	194.1173	0.8665	1.0204	1.0133	0.8288	0.9271	1.0063	0.974	1.064	1.0822	1.0516	1.085
9	Love Drug	180	506.4	505.6407	180.0866	1.324	1.5555	1.5448	1.2935	1.303	1.3073	1.4168	1.4861	1.4673	1.7218	1.4975
10	Ketamine	238	600	599.9313	238.0988	1.1623	1.1566	1.1886	1.1839	1.259	1.2549	1.2607	1.2335	1.2798	1.2108	1.2105
11	FM2	314	1044	1044.274	314.093	0.9901	0.9353	0.9865	0.9695	0.9464	0.9406	0.9512	0.9684	0.9777	0.9584	0.9852
12	Nimetazepam	296	1041	1040.522	296.1024	0.9651	0.9587	0.9532	0.9341	0.9048	0.9264	0.9336	0.9345	0.9144	0.9565	0.9393

Peak analysis (targeted)

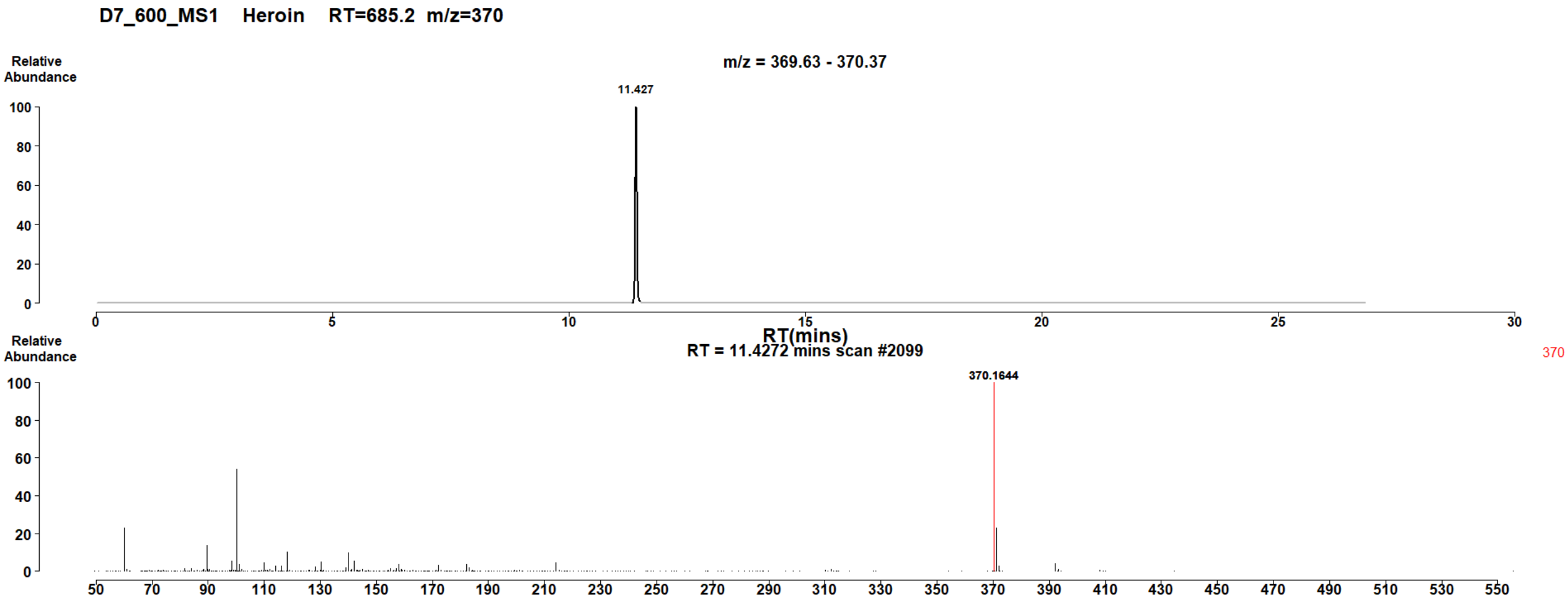
Output file

S/N (MS2)

Peak_Index	Name	parent_ion	mz		Ret_Time.sec	Ret_Time_est.sec	mz_est	D1_50_MSMS	D10_900_MSMS	D11_1000_MSMS	D2_100_MSMS	D3_200_MSMS	D4_300_MSMS	D5_400_MSMS	D6_500_MSMS	D7_600_MSMS	D8_700_MSMS	D9_800_MSMS
1	Heroin	370	370	NA	685.5322	370.1636	2.2754	2.1575	2.2095	1.7426	1.5182	1.5774	1.8792	1.919	1.9821	1.9042	2.0616	
2	Heroin	370	328	NA	685.5322	328.1531	2.2624	2.1629	2.2075	1.7496	1.5132	1.5627	1.8573	1.9065	1.9546	1.8699	2.0653	
3	Heroin	370	268	NA	685.5322	268.1322	2.2613	2.1554	2.2105	1.7726	1.5105	1.57	1.8721	1.9139	1.9731	1.8614	2.0704	
4	Heroin	370	211	NA	685.5322	211.0747	2.2738	2.1484	2.2039	1.7586	1.5087	1.5658	1.8721	1.9124	1.9829	1.8725	2.0448	
5	Heroin	370	193	NA	685.5322	193.0641	2.2736	2.1775	2.2153	1.7398	1.5218	1.5767	1.8779	1.8741	1.9691	1.8508	2.0683	
6	Heroin	370	58	NA	686.438	57.99157	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
7	Morphine	286	286	NA	145.5	286.143	1.5084	1.498	1.4622	1.6833	1.794	1.6962	1.7063	1.6977	1.6228	1.5853	1.5373	
8	Morphine	286	268	NA	145.5	268.1305	1.4903	1.5157	1.4851	1.7056	1.8104	1.7309	1.7479	1.7242	1.6543	1.6184	1.5653	
9	Morphine	286	229	NA	145.5	229.0854	1.5004	1.5458	1.506	1.7119	1.7769	1.7119	1.7187	1.6686	1.6353	1.6151	1.561	
10	Morphine	286	201	NA	145.5	201.0879	1.48	1.5655	1.5007	1.7254	1.7964	1.7166	1.7216	1.6632	1.6224	1.5915	1.5471	
11	Cocaine	304	304	NA	700.1591	304.1536	1.9155	1.4562	1.4427	2.0541	1.9777	1.8857	1.7429	1.6952	1.6246	1.5651	1.4984	
12	Cocaine	304	182	NA	699.8602	182.117	1.8864	1.5056	1.4639	2.0907	2.0307	1.9291	1.8093	1.7527	1.6855	1.6817	1.5992	
13	Cocaine	304	150	NA	699.8602	150.091	1.8077	1.5238	1.4723	2.0971	2.0433	1.9661	1.8037	1.7921	1.6852	1.6823	1.5917	
14	Cocaine	304	105	NA	699.8602	105.0332	1.7233	1.6279	1.5504	2.1	2.0632	1.9639	1.8819	1.8348	1.7372	1.7087	1.6749	
15	Cocaine	304	82	NA	699.8602	82.0649	1.7085	1.623	1.5539	2.1039	2.0543	1.97	1.8798	1.8226	1.7479	1.7198	1.6732	

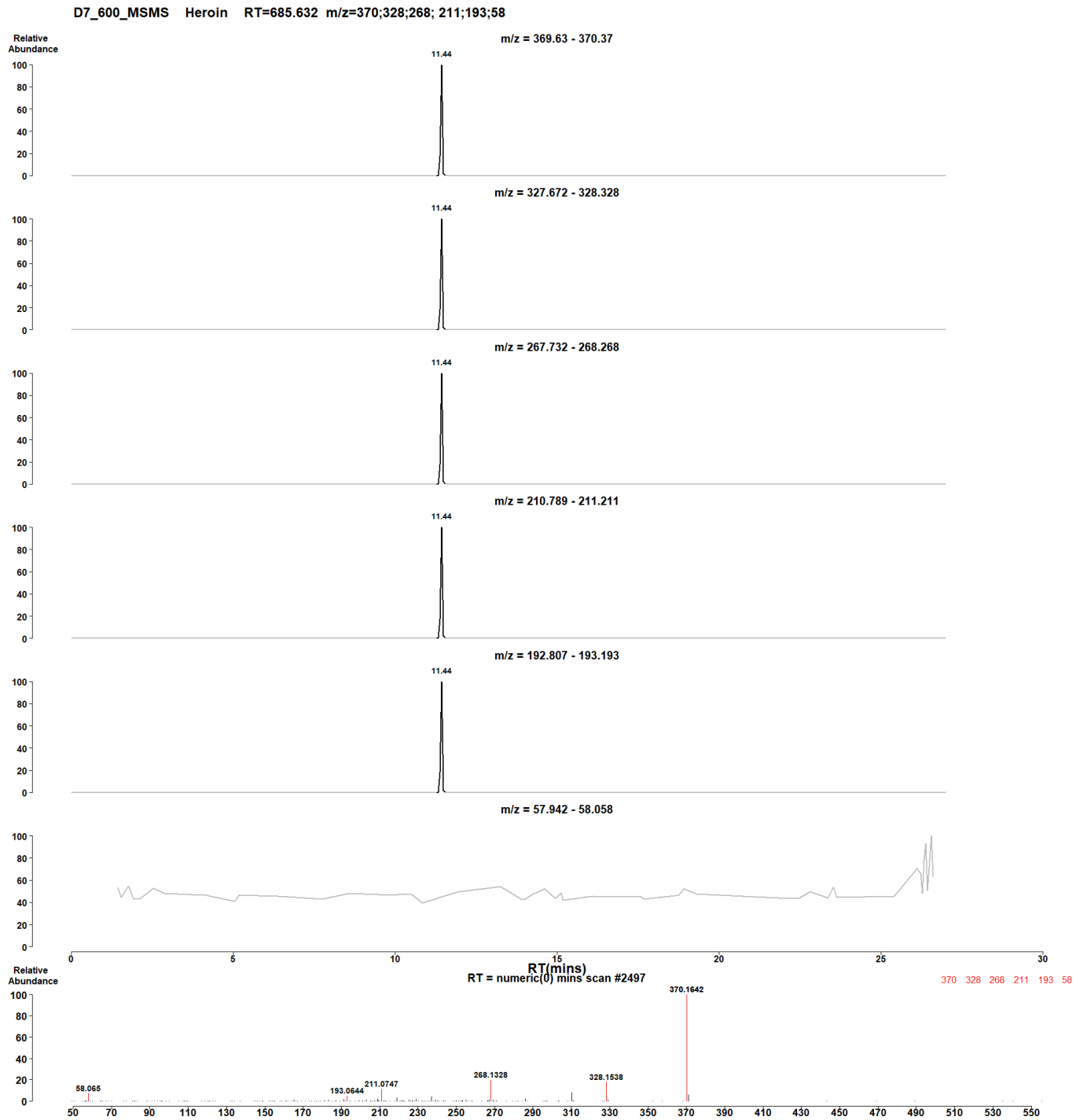
Peak analysis (targeted)

Figure (MS1)

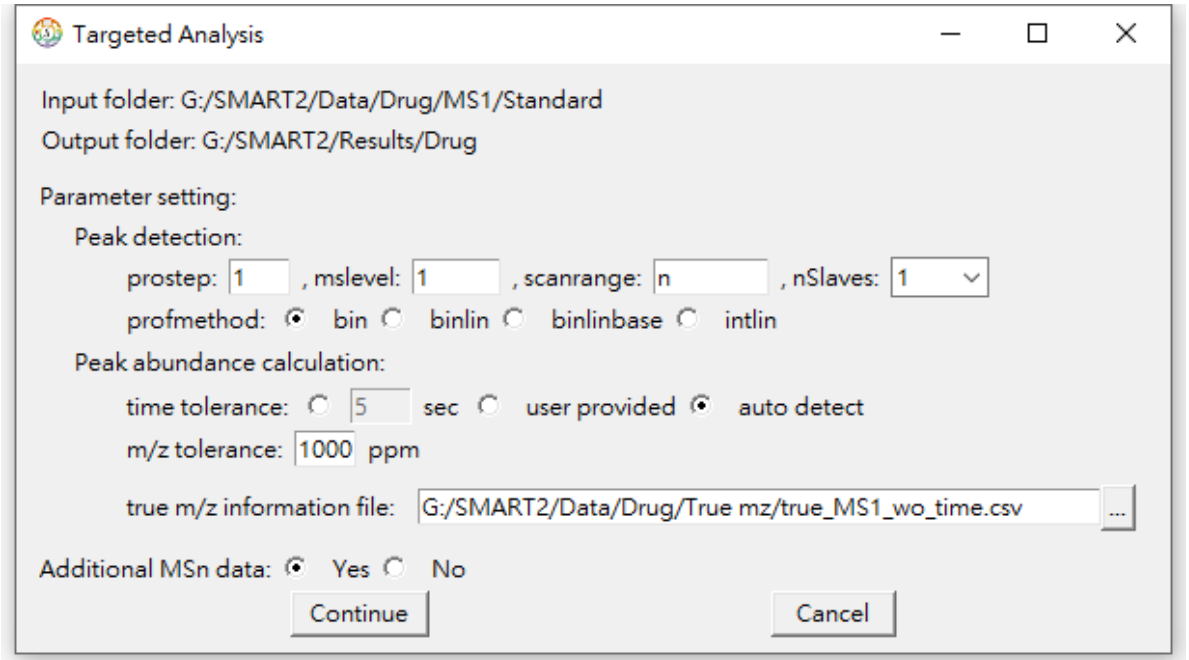


Peak analysis (targeted)

Figure (MS2)

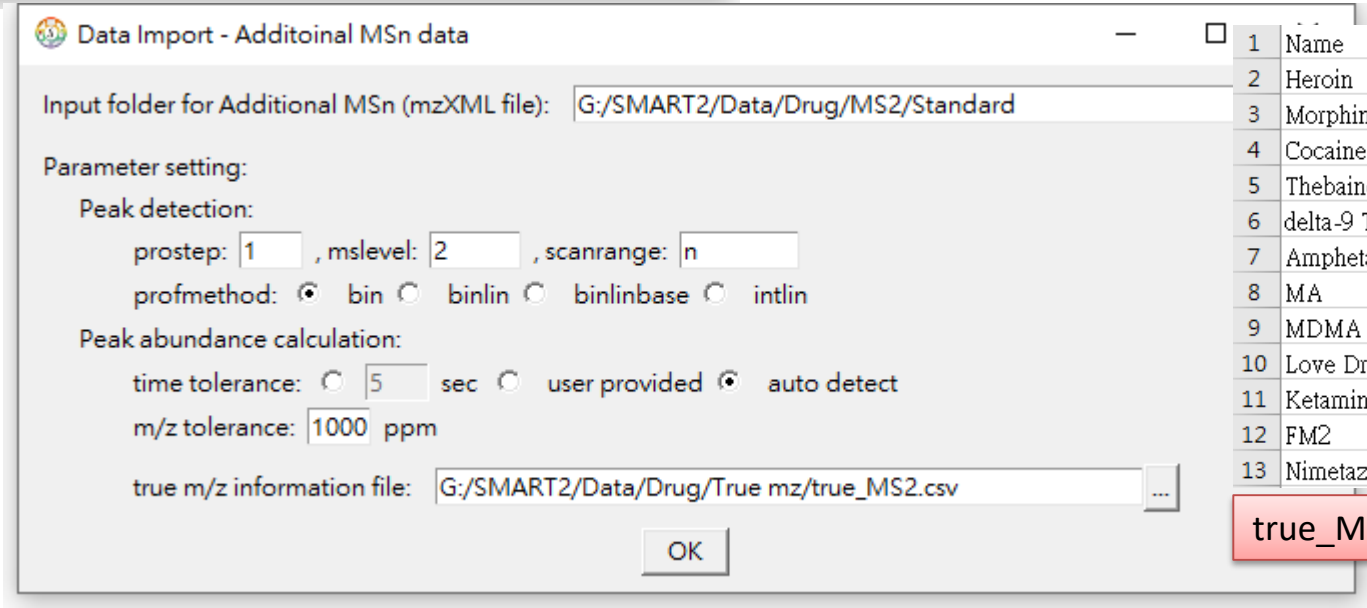


Auto detect



	Name	mz	Ret_Time.sec
2	Heroin	370	NA
3	Morphine	286	NA
4	Cocaine	304	NA
5	Thebaine	312	NA
6	delta-9 THC	315	NA
7	Amphetamine	136	NA
8	MA	150	NA
9	MDMA	194	NA
10	Love Drug	180	NA
11	Ketamine	238	NA
12	FM2	314	NA
13	Nimetazepam	296	NA

true_MS1_wo_time.csv

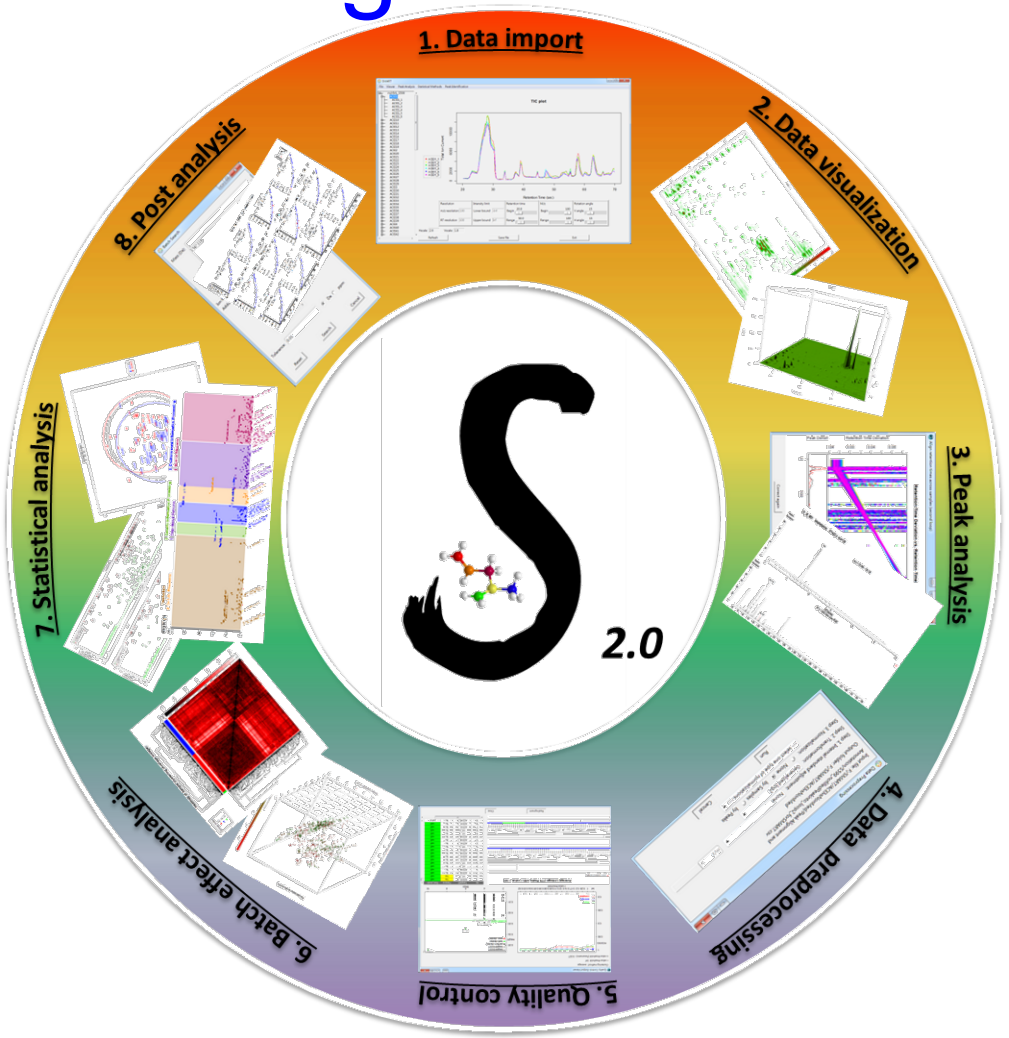


	Name	parent_ion	daughter_ion
2	Heroin	370	370;328;268; 211;193;58
3	Morphine	286	286; 268; 229; 201
4	Cocaine	304	304; 182; 150; 105; 82
5	Thebaine	312	312; 281; 266; 251; 221; 58
6	delta-9 THC	315	315; 259; 193; 135; 123; 107; 93
7	Amphetamine	136	136; 119; 91
8	MA	150	150; 119; 91
9	MDMA	194	194; 163; 135; 133; 105
10	Love Drug	180	180; 163; 135; 133; 105
11	Ketamine	238	238; 220; 179; 163; 152; 125
12	FM2	314	314; 200; 286; 268
13	Nimetazepam	296	296; 268; 250; 222; 193; 165

true_MS2.csv

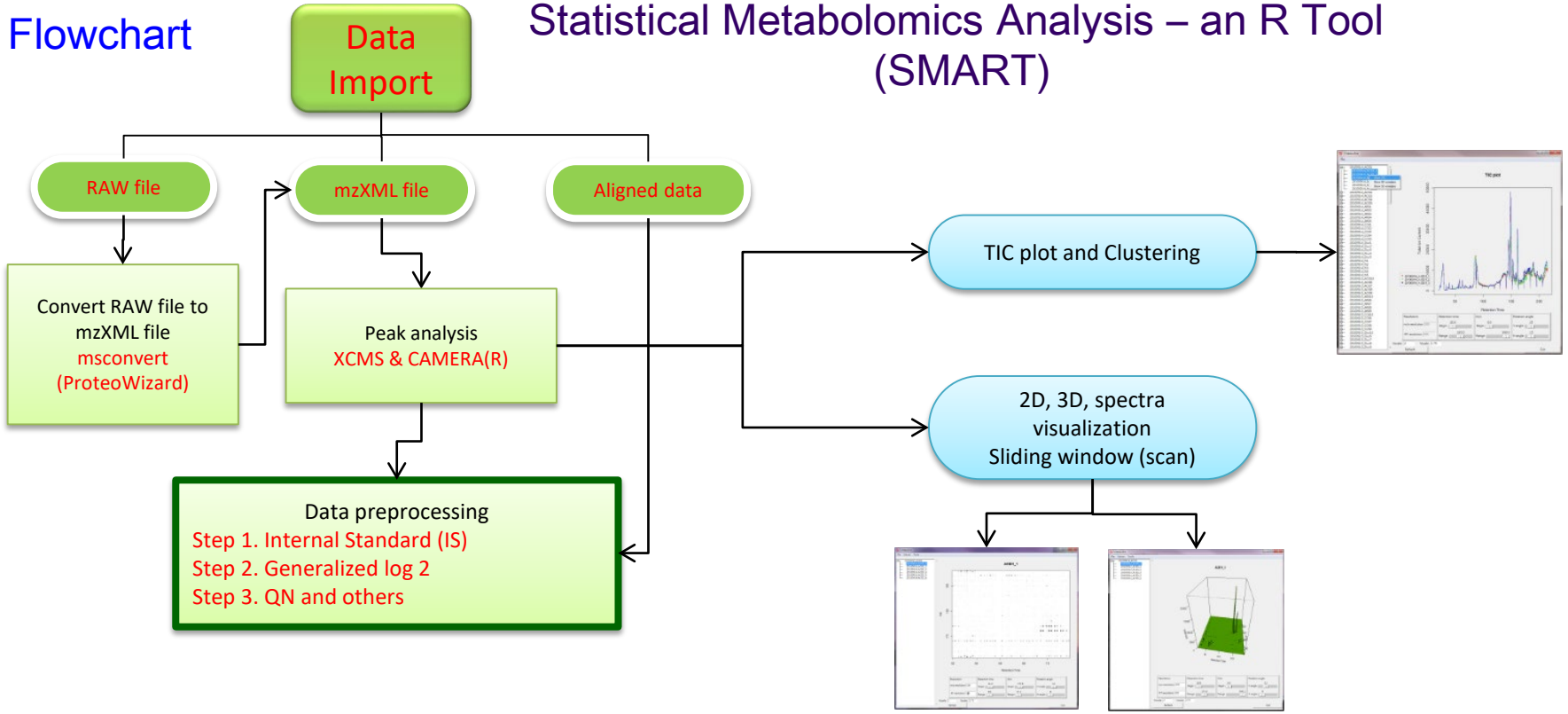
Statistical Metabolomics Analysis – an R Tool

4. Data preprocessing



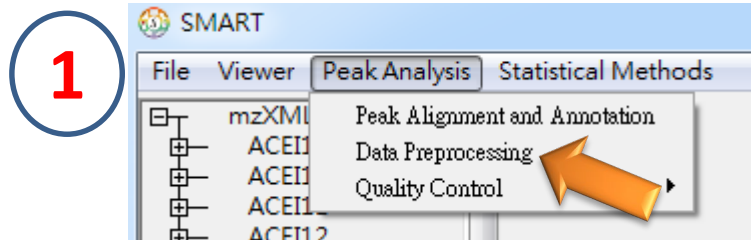
Statistical Metabolomics Analysis – an R Tool (SMART)

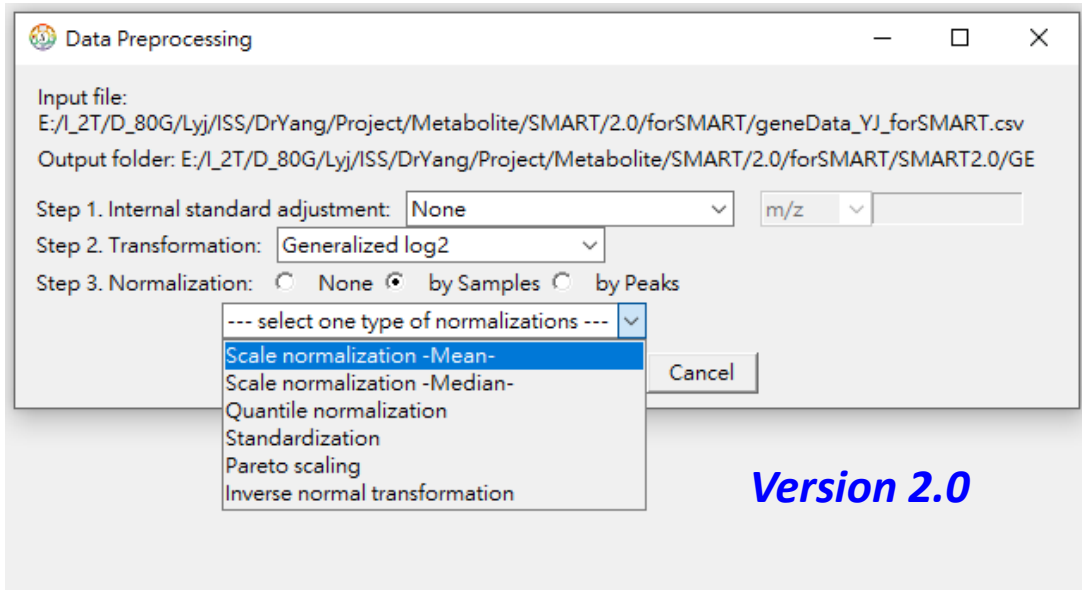
Flowchart

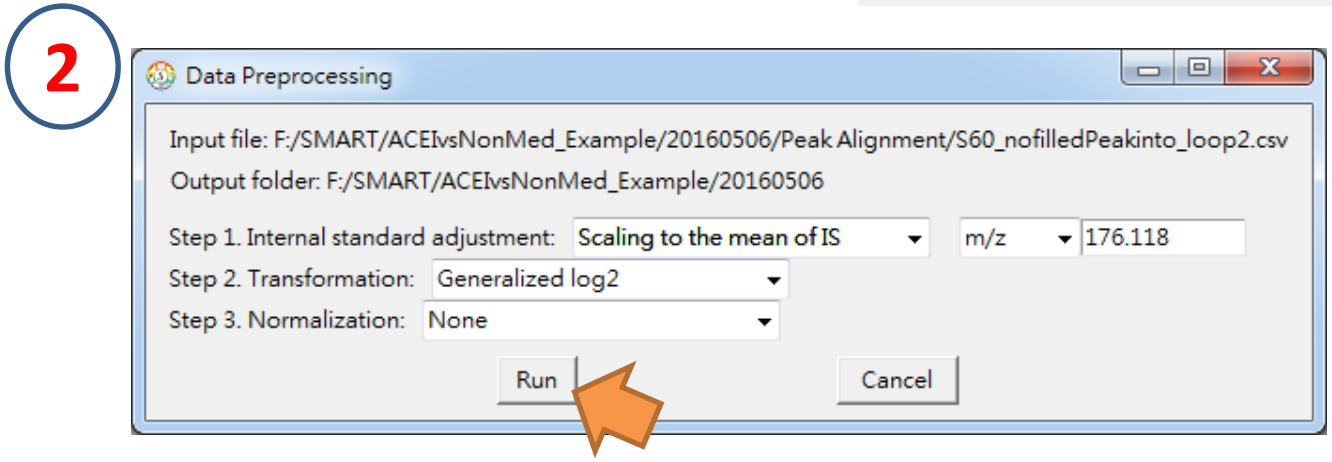


Statistical Metabolomics Analysis - an R Tool

- 4. Data preprocessing



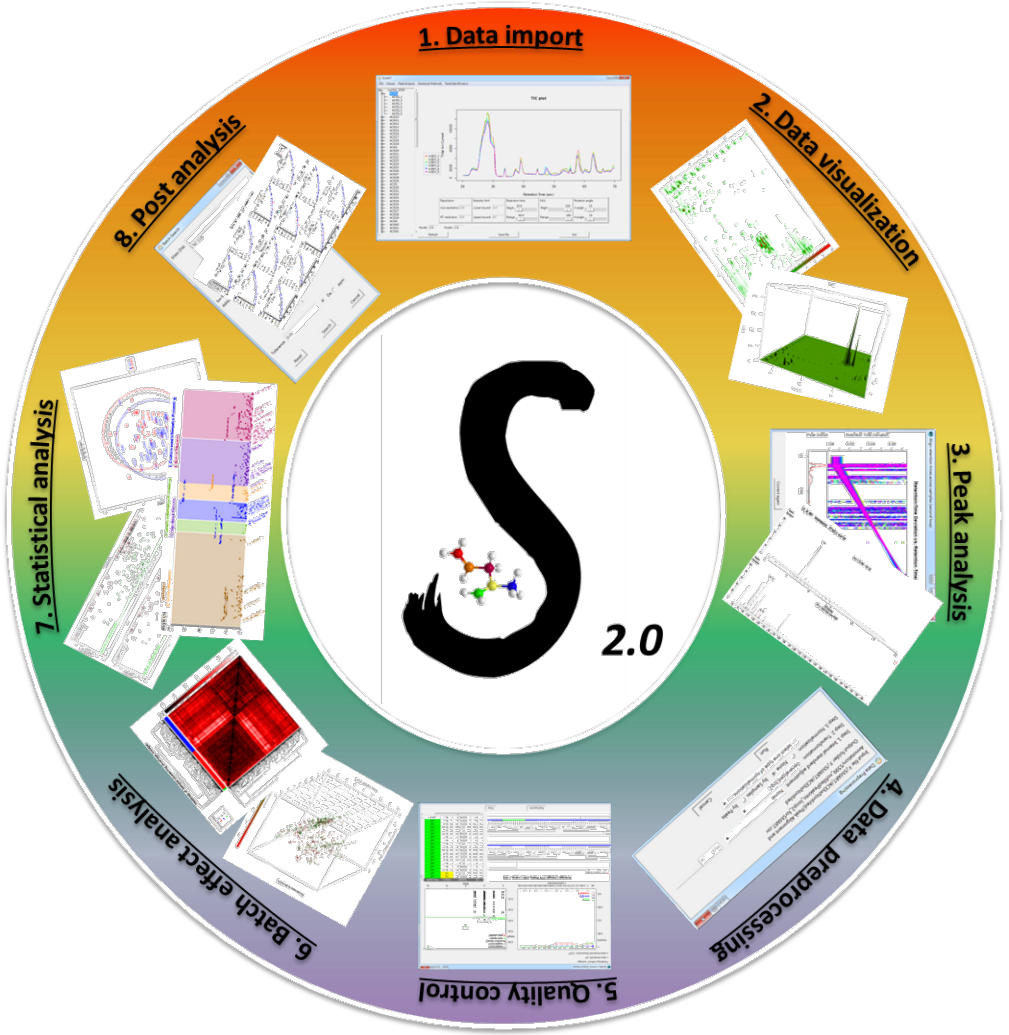




Data preprocessing

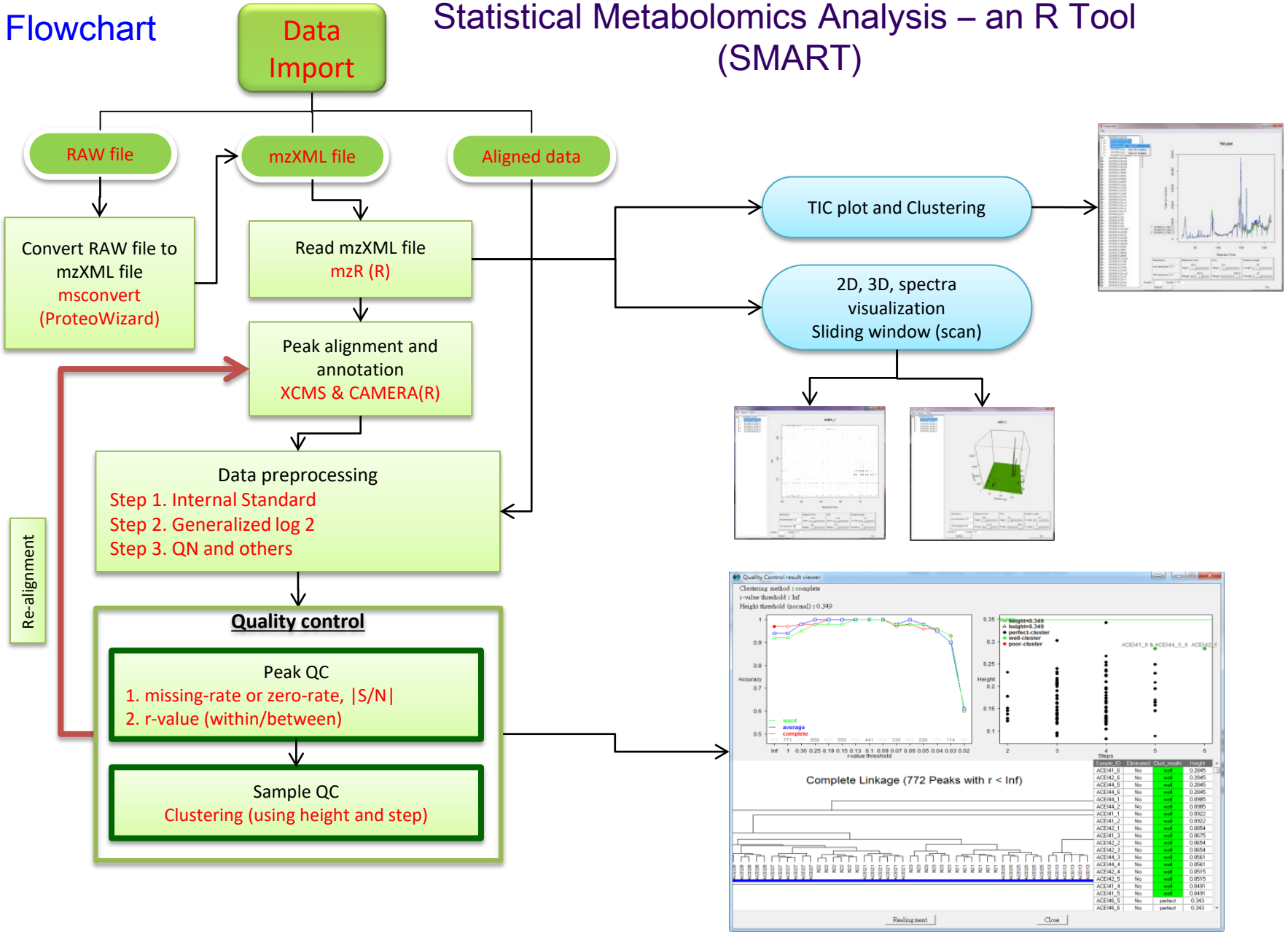
Statistical Metabolomics Analysis – an R Tool

5. Quality control

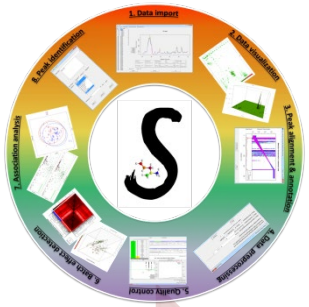


Flowchart

Statistical Metabolomics Analysis – an R Tool (SMART)



5. Quality control -- sample filtering



1. Cluster analysis based on good-quality peaks

- Quality index for peaks : r-value = SSW / SSB
- Average-linkage , complete-linkage , or Ward's method

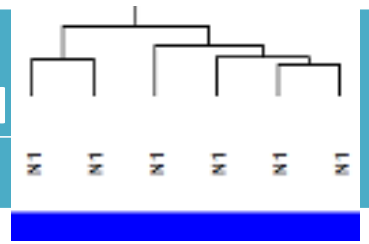
2. Distance measurement (h)

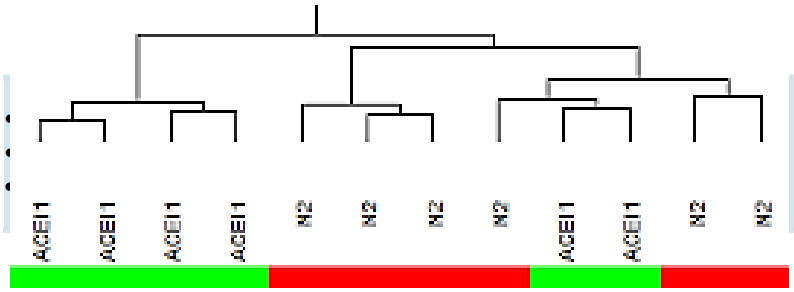
- h = log((1 - p) / 2) , p = Pearson's correlation coefficient

3. Subjects categorization

- perfectly-clustered : all replicate samples of this subject are clustered together
- well-clustered : at least 2/3 replicate samples of this subject are clustered together
- poorly-clustered : otherwise

4. s-val





perfectly-clustered

6 replicates

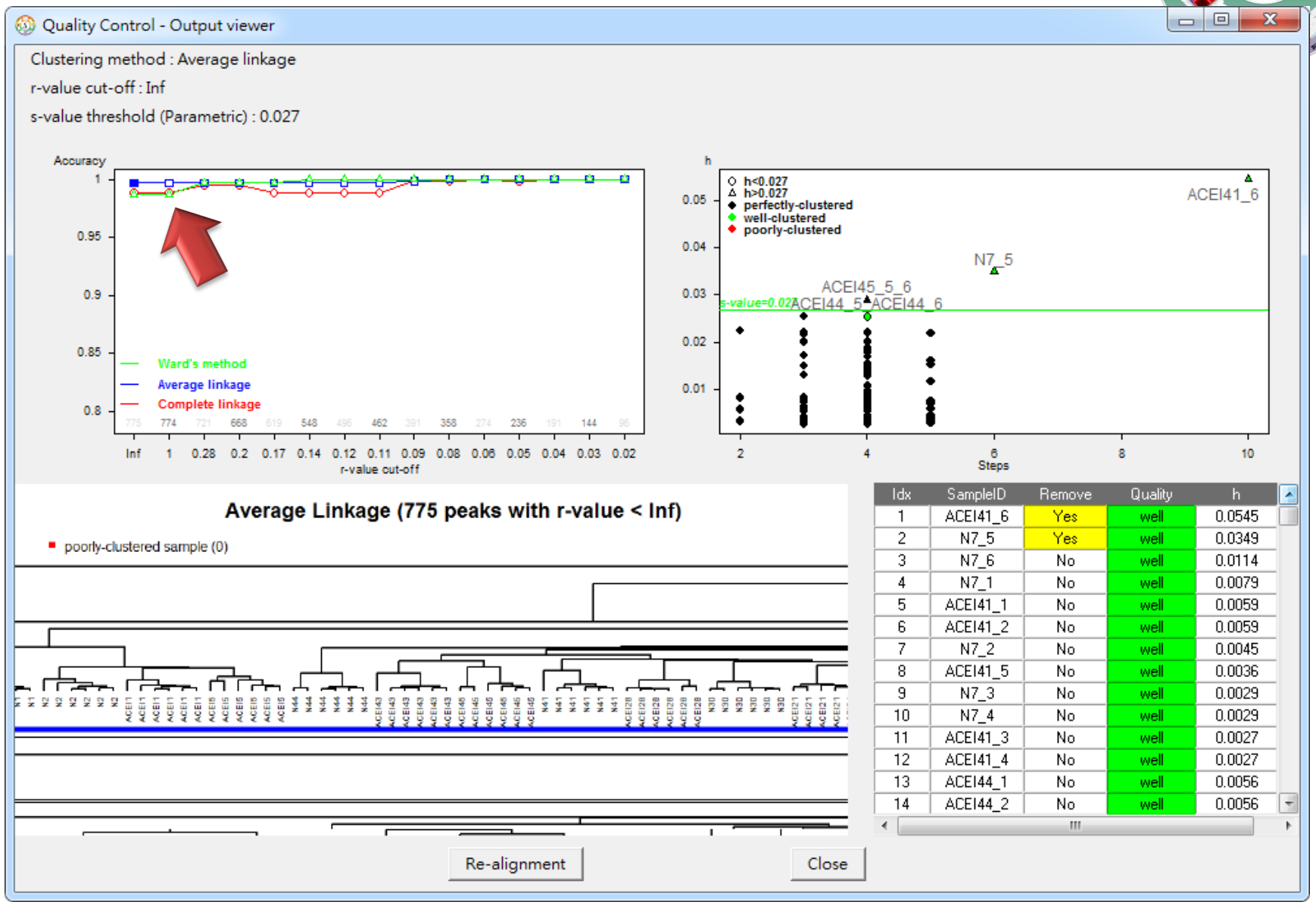
well-clustered

4 or 5 replicates

poorly-clustered

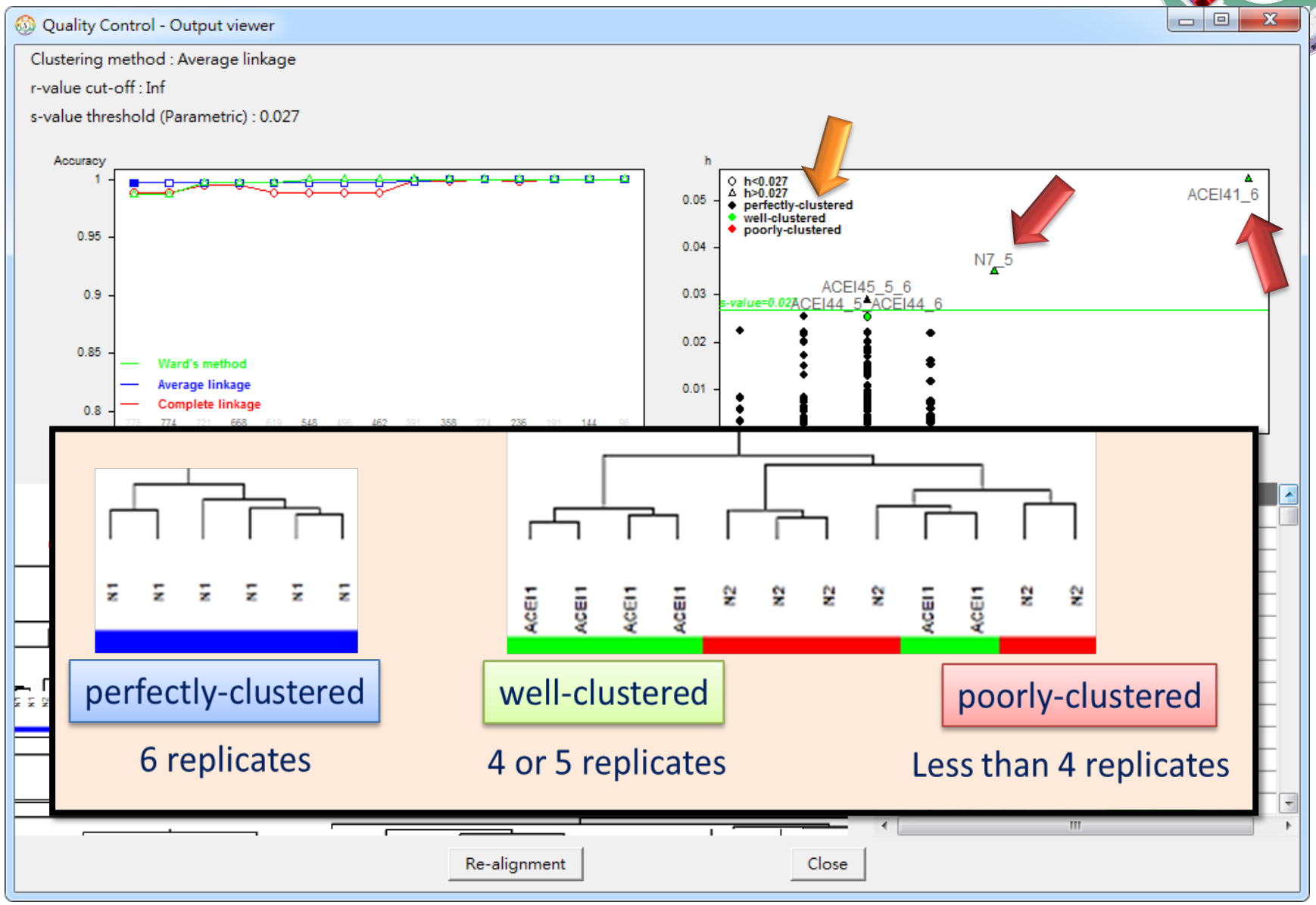
Less than 4 replicates

5. Quality control -- sample filtering (2)

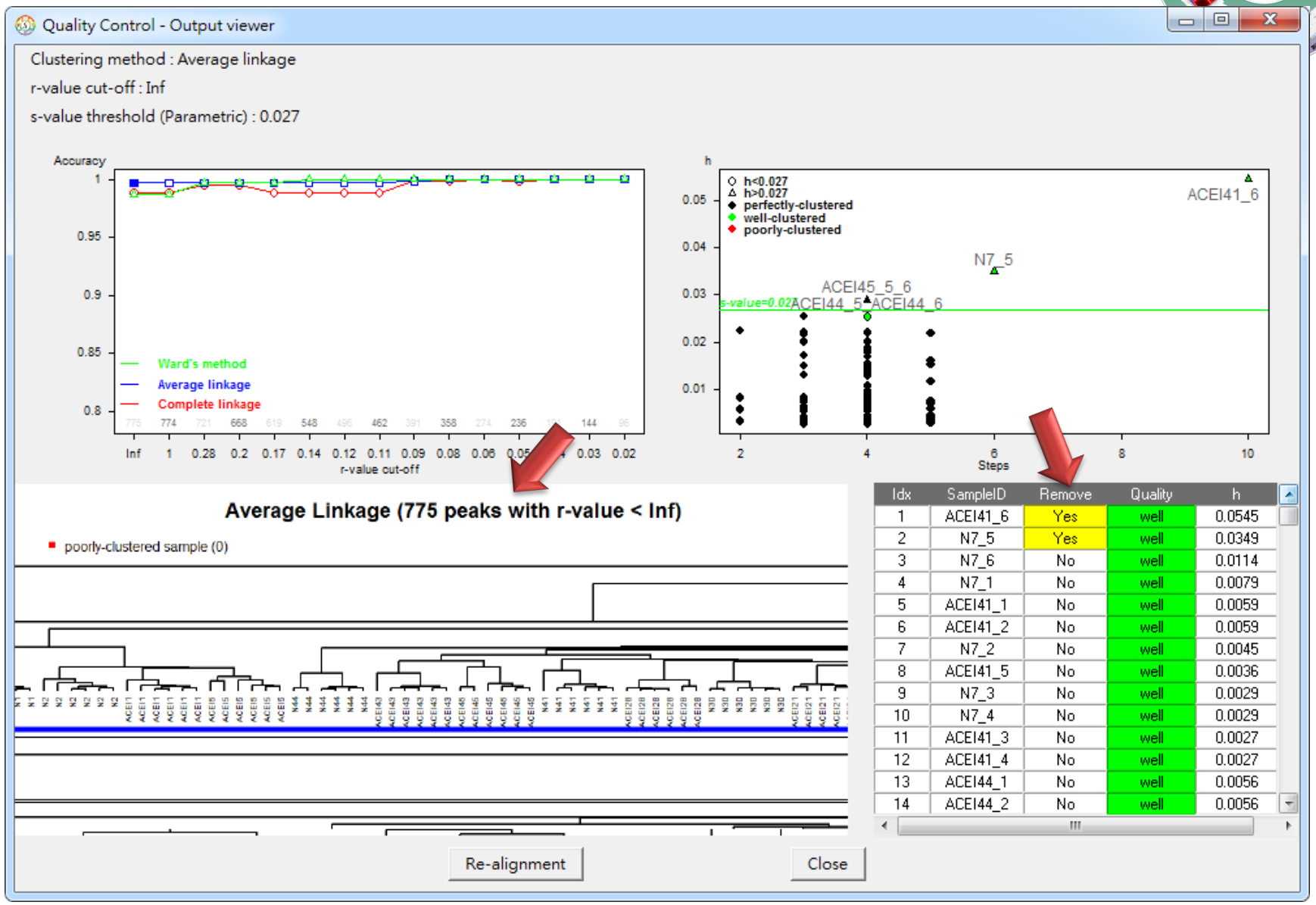


Idx	SampleID	Remove	Quality	h
1	ACEI41_6	Yes	well	0.0545
2	N7_5	Yes	well	0.0349
3	N7_6	No	well	0.0114
4	N7_1	No	well	0.0079
5	ACEI41_1	No	well	0.0059
6	ACEI41_2	No	well	0.0059
7	N7_2	No	well	0.0045
8	ACEI41_5	No	well	0.0036
9	N7_3	No	well	0.0029
10	N7_4	No	well	0.0029
11	ACEI41_3	No	well	0.0027
12	ACEI41_4	No	well	0.0027
13	ACEI44_1	No	well	0.0056
14	ACEI44_2	No	well	0.0056

5. Quality control -- sample filtering (2)



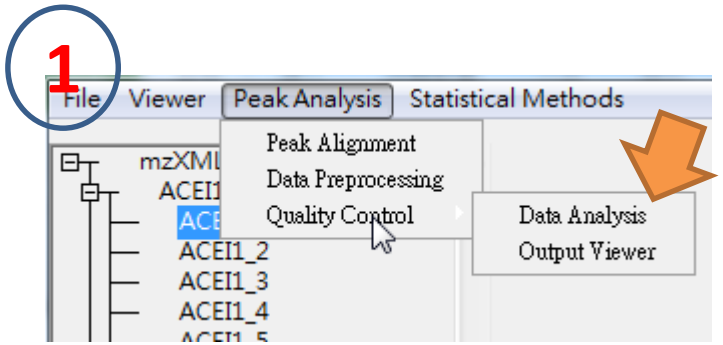
5. Quality control -- sample filtering (2)

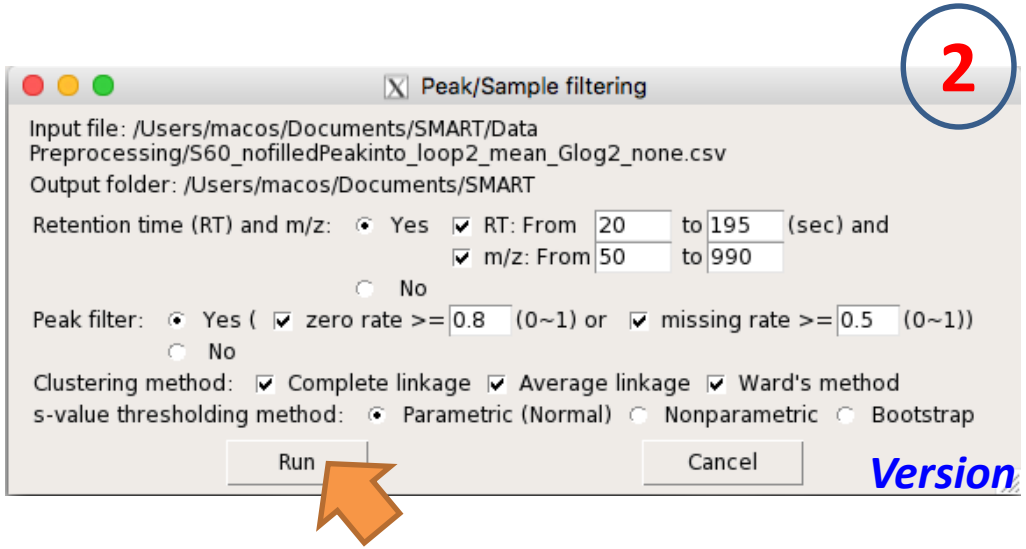


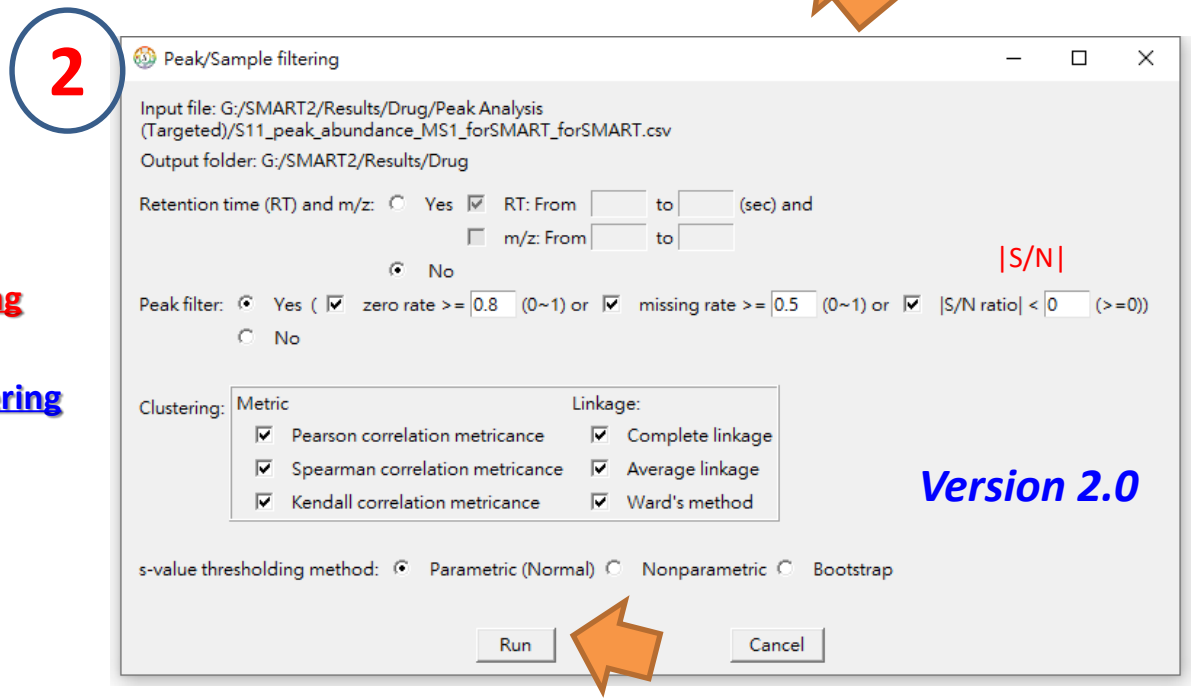
Idx	SampleID	Remove	Quality	h
1	ACEI41_6	Yes	well	0.0545
2	N7_5	Yes	well	0.0349
3	N7_6	No	well	0.0114
4	N7_1	No	well	0.0079
5	ACEI41_1	No	well	0.0059
6	ACEI41_2	No	well	0.0059
7	N7_2	No	well	0.0045
8	ACEI41_5	No	well	0.0036
9	N7_3	No	well	0.0029
10	N7_4	No	well	0.0029
11	ACEI41_3	No	well	0.0027
12	ACEI41_4	No	well	0.0027
13	ACEI44_1	No	well	0.0056
14	ACEI44_2	No	well	0.0056

Statistical Metabolomics Analysis – an R Tool

- 5. Quality control





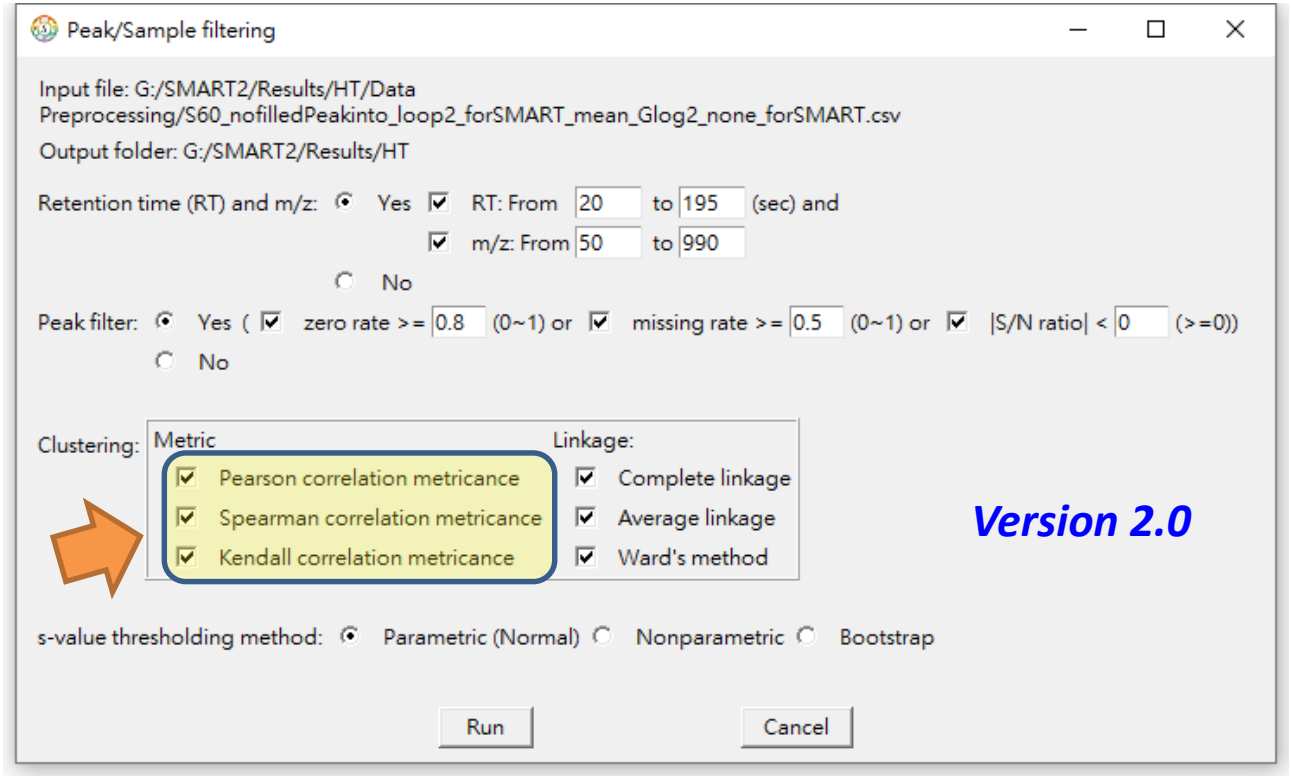


Peak filtering

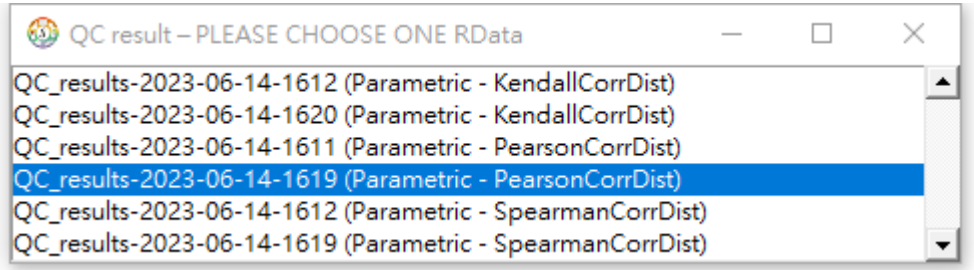
Sample filtering

Statistical Metabolomics Analysis – an R Tool

- 5. Quality control



3

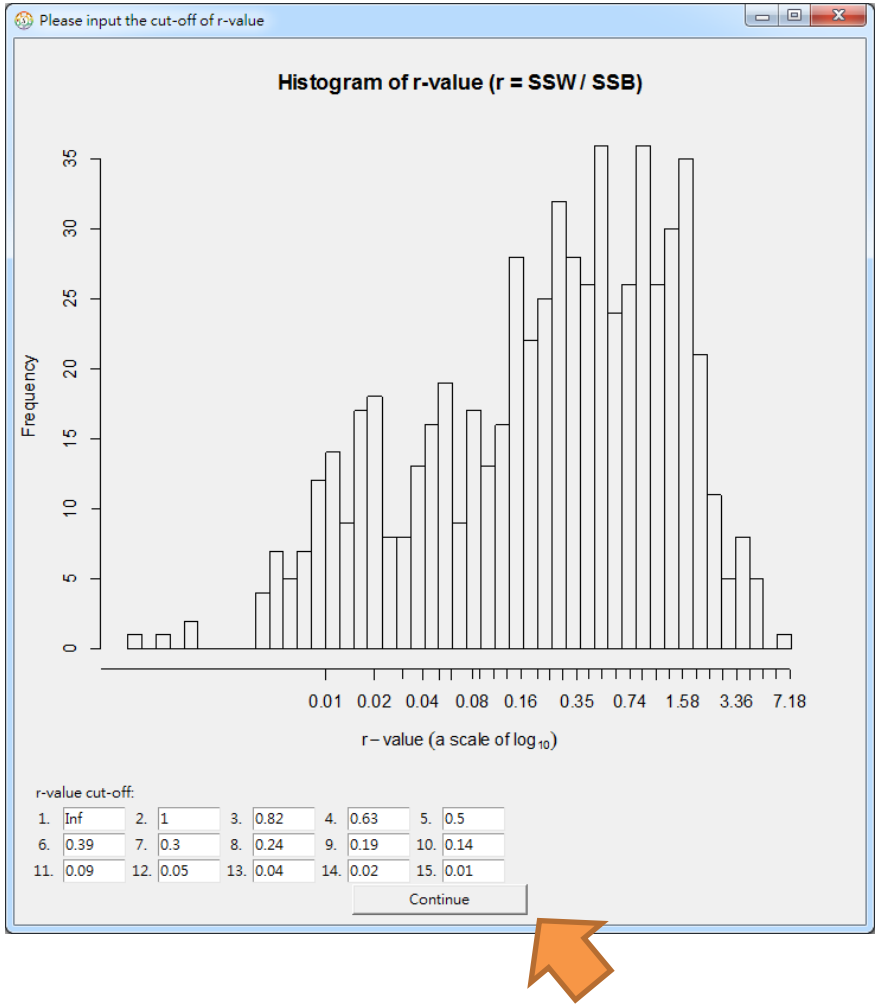


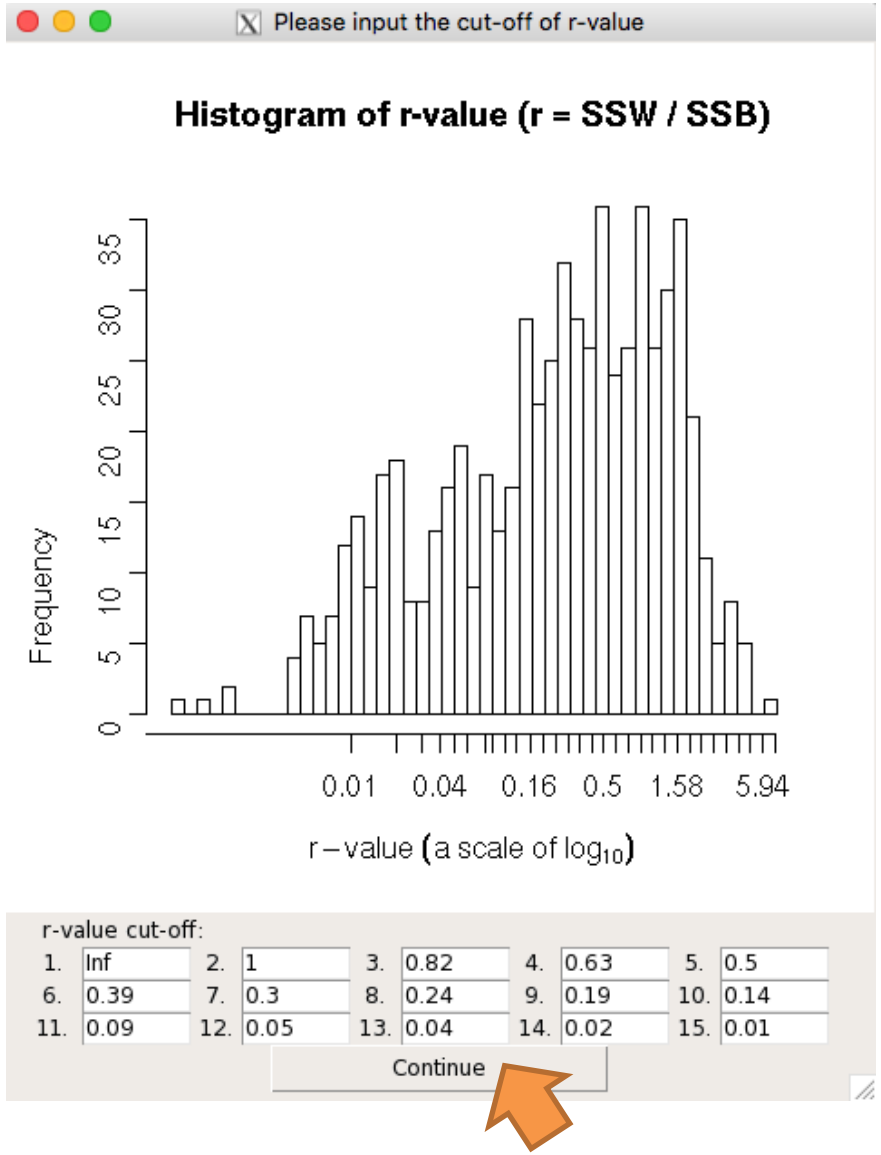
Statistical Metabolomics Analysis – an R Tool

- 5. Quality control

Statistical Metabolomics Analysis – an R Tool

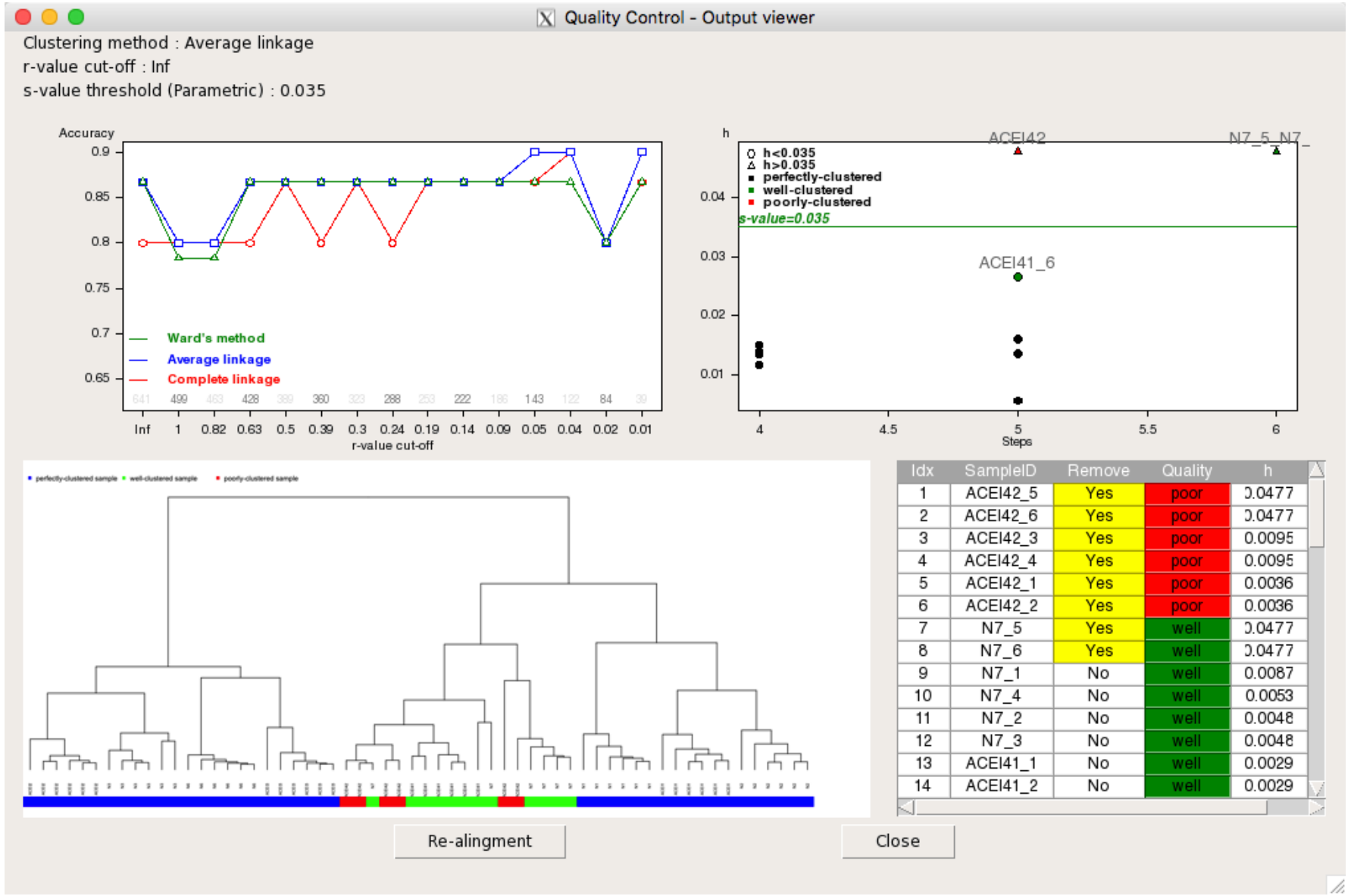
- 5. Quality control





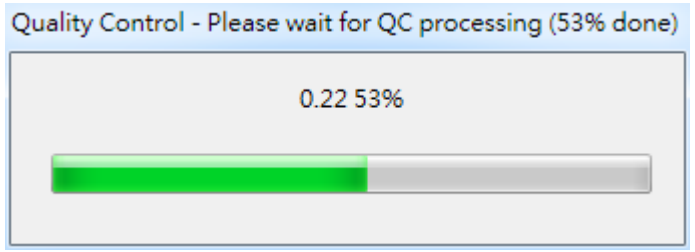
Statistical Metabolomics Analysis – an R Tool

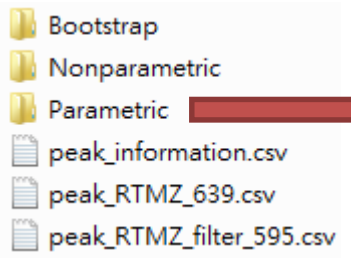
- 5. Quality control

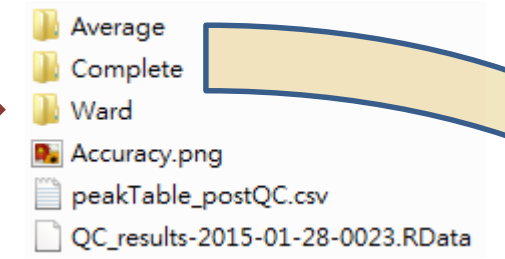


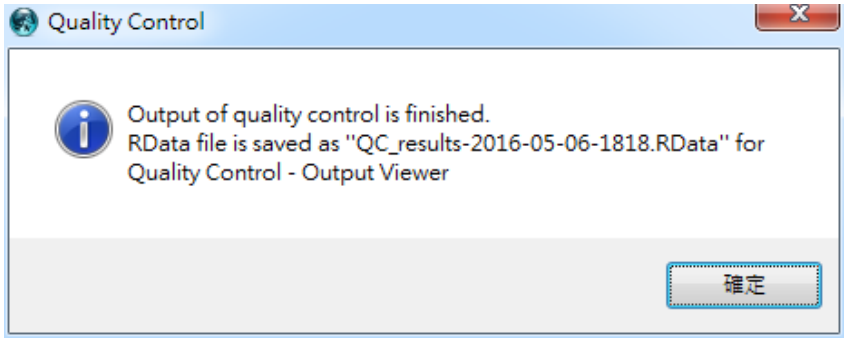
Statistical Metabolomics Analysis – an R Tool

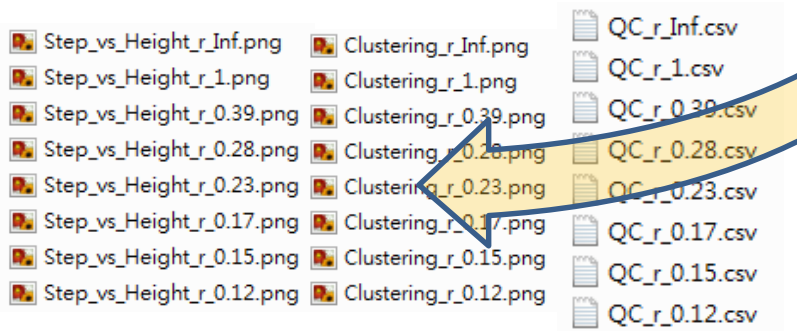
- 5. Quality control

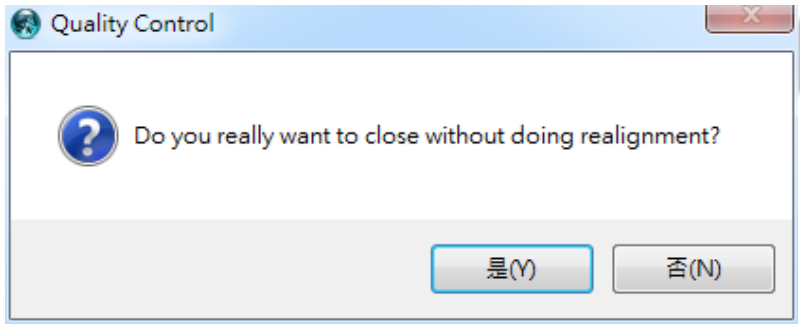










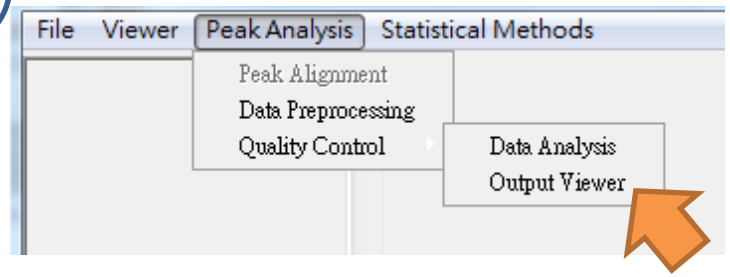


Output path: *outpath*\Quality Control\Normal\Average

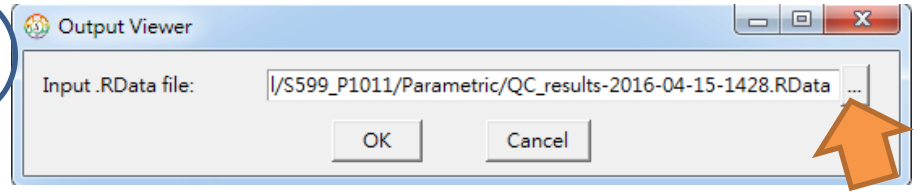
Statistical Metabolomics Analysis – an R Tool

- 5. Quality control – output viewer

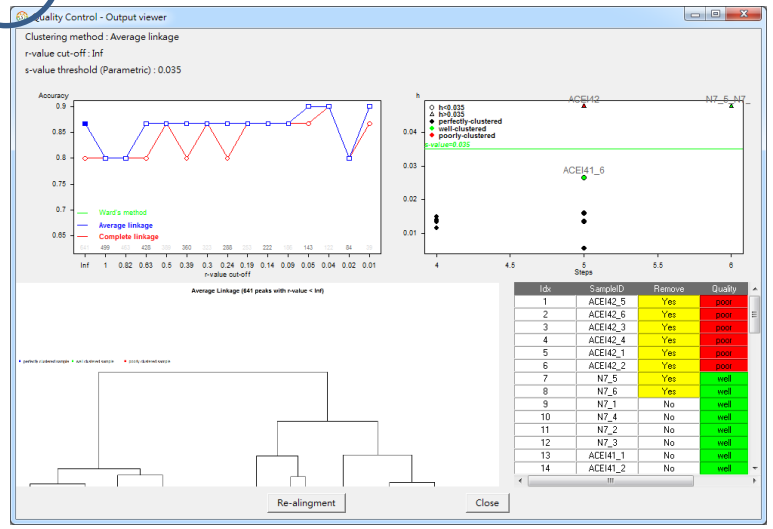
1



2



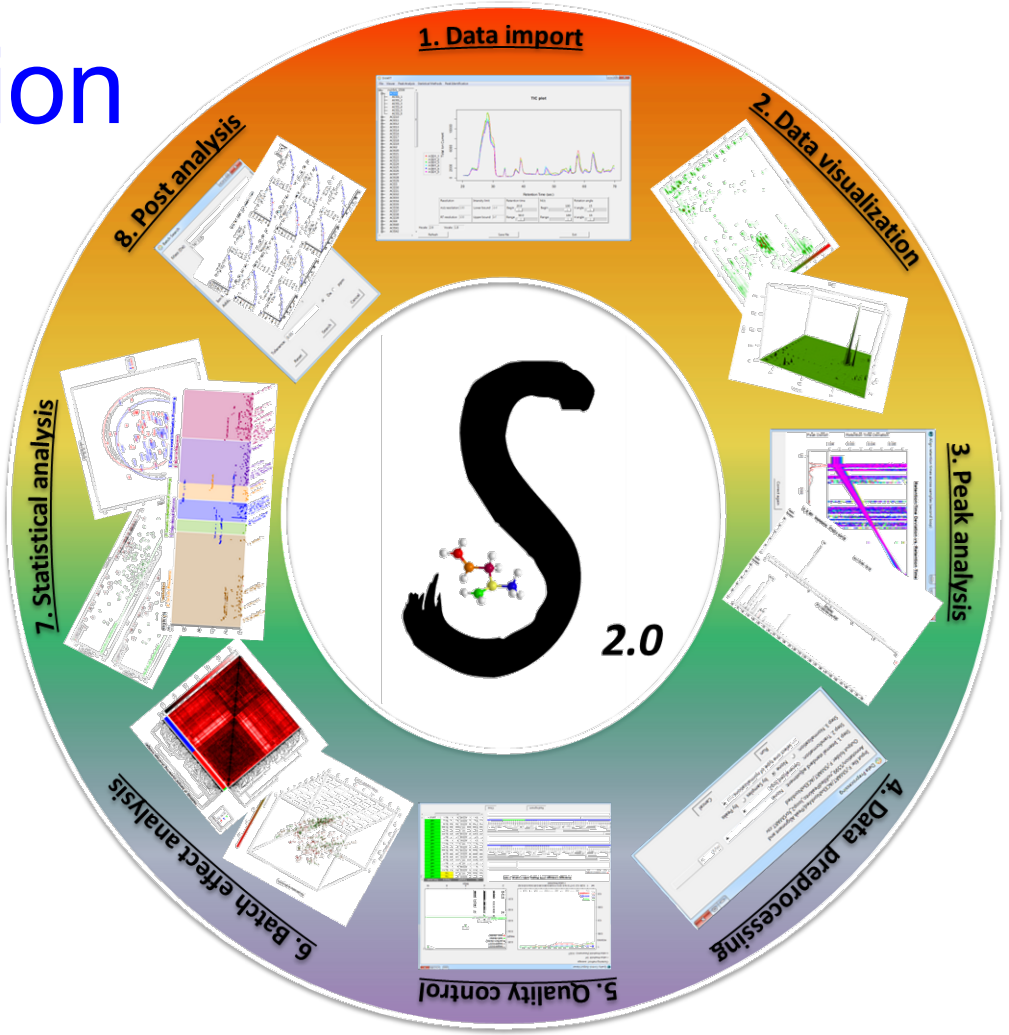
3



Quality control (1st)

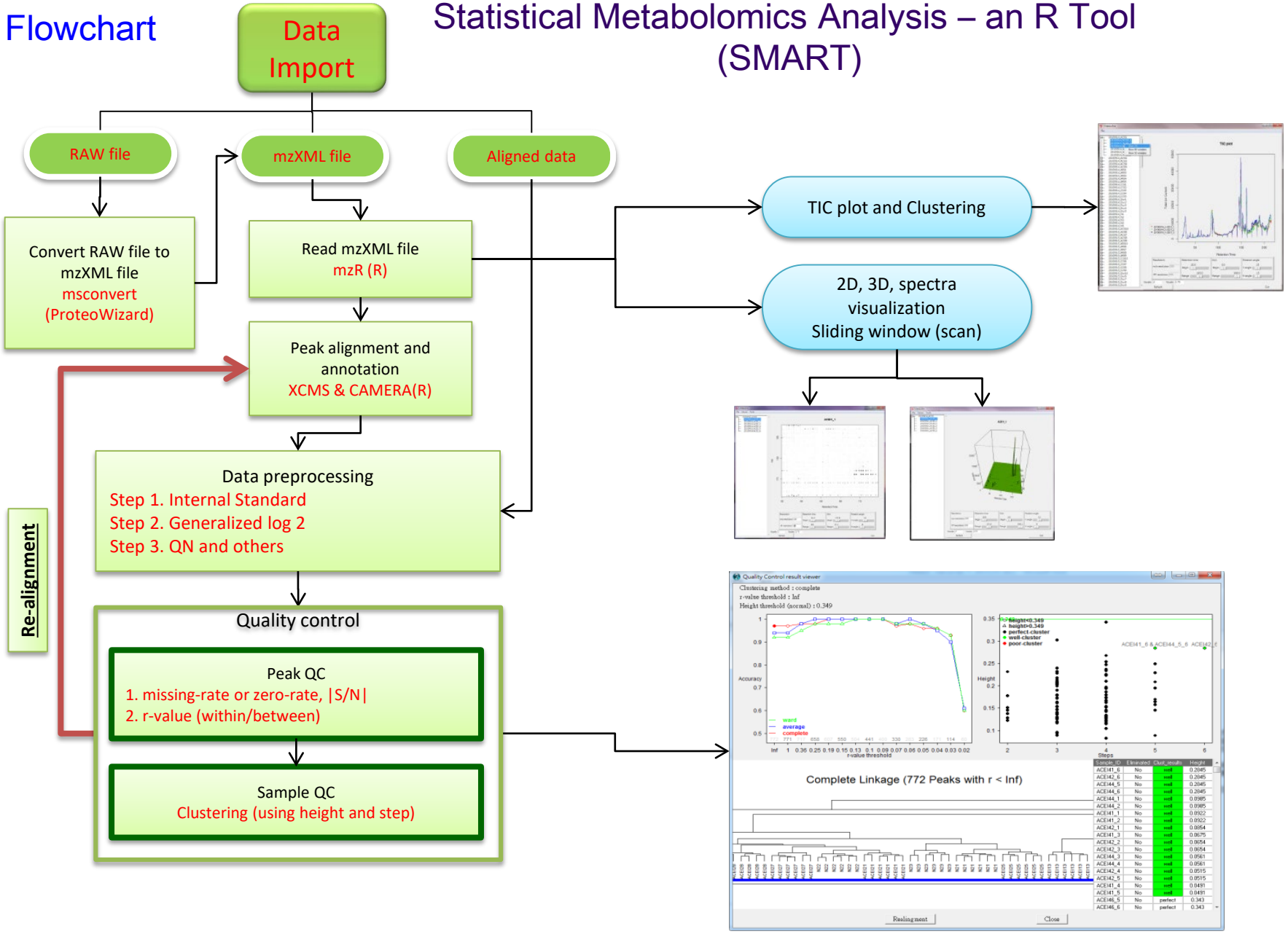
Statistical Metabolomics Analysis – an R Tool

Peak re-alignment and annotation



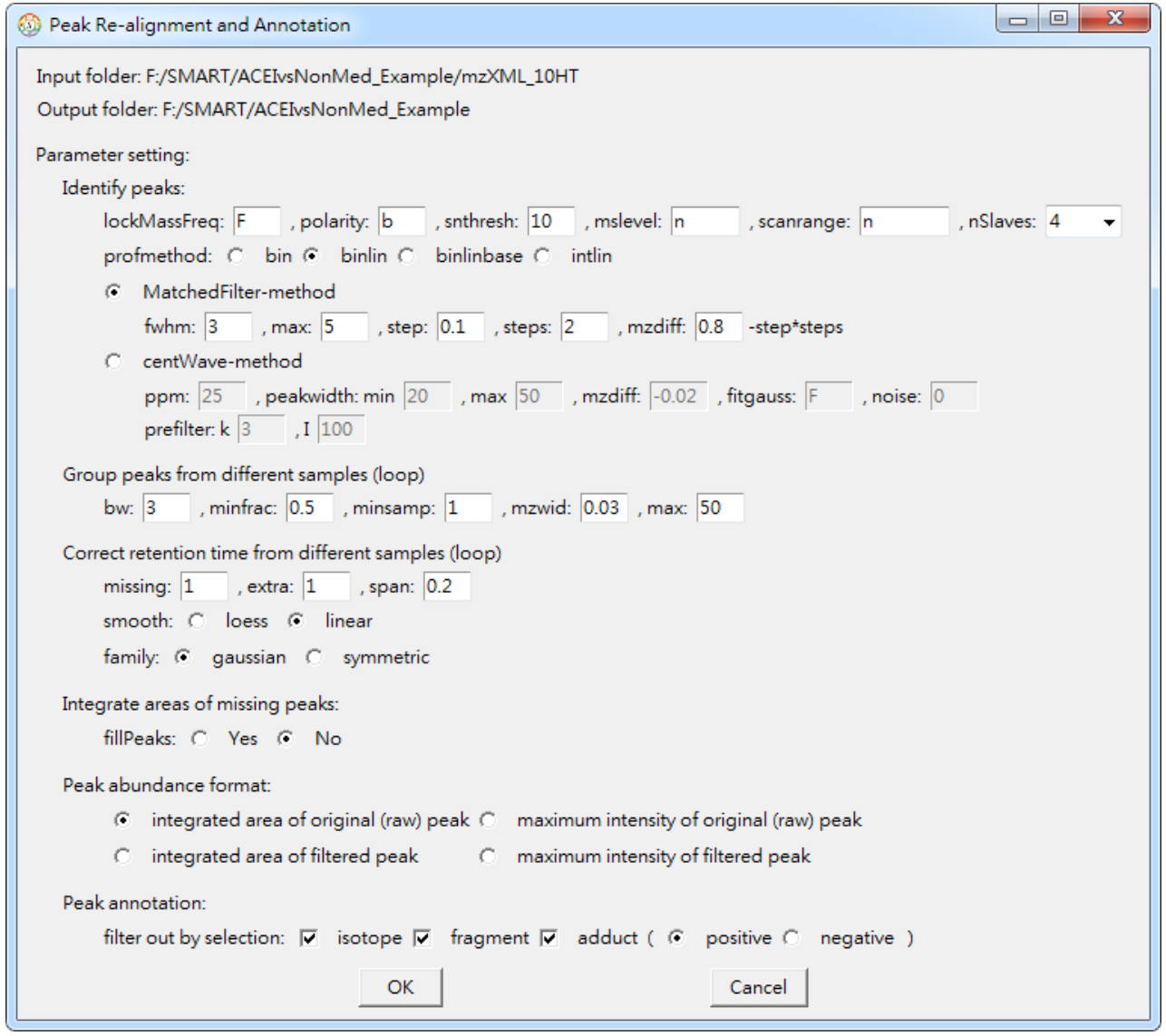
Statistical Metabolomics Analysis – an R Tool (SMART)

Flowchart



Re-alignment & annotation

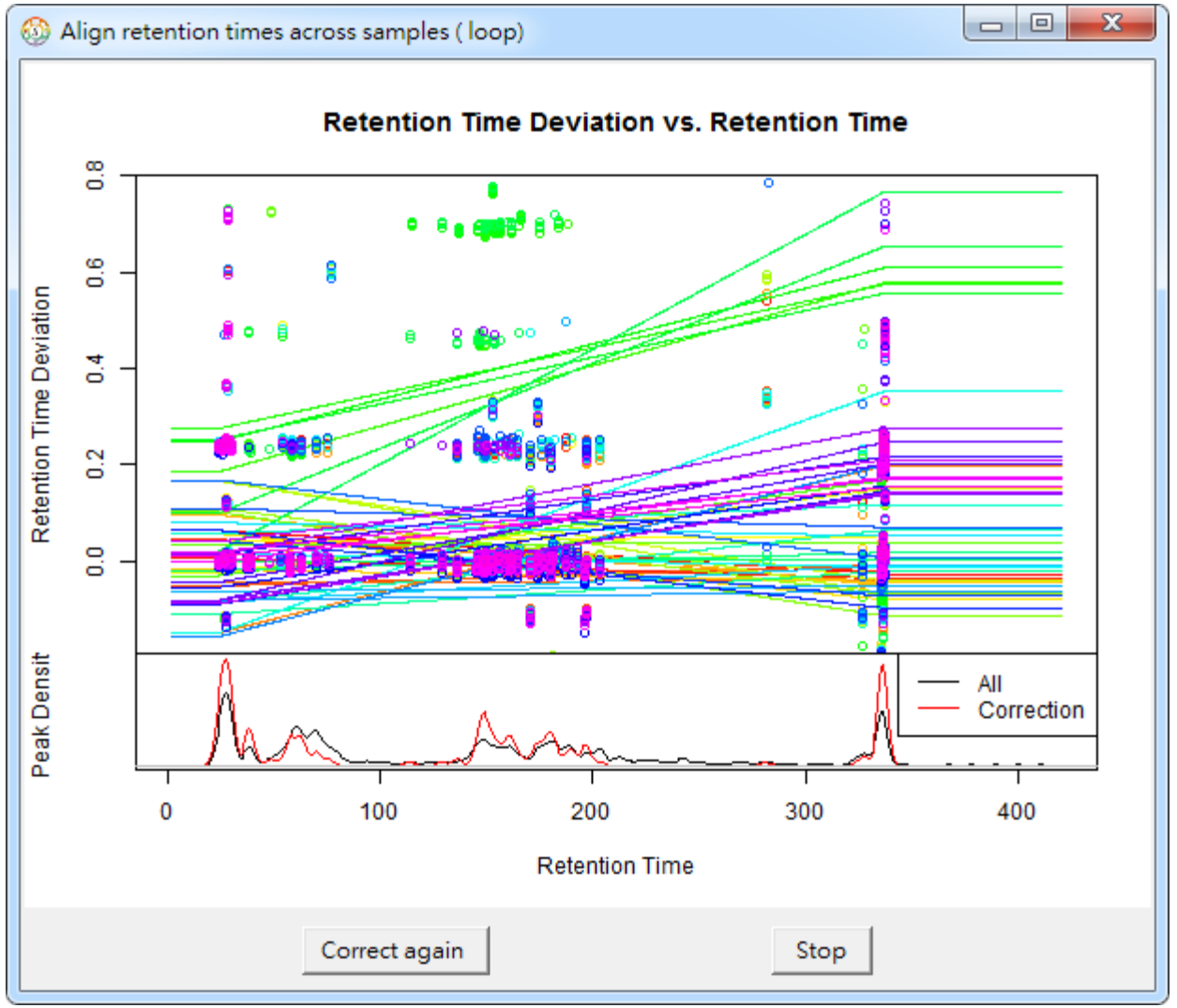
- Re-alignment & annotation



Statistical Metabolomics Analysis – an R Tool

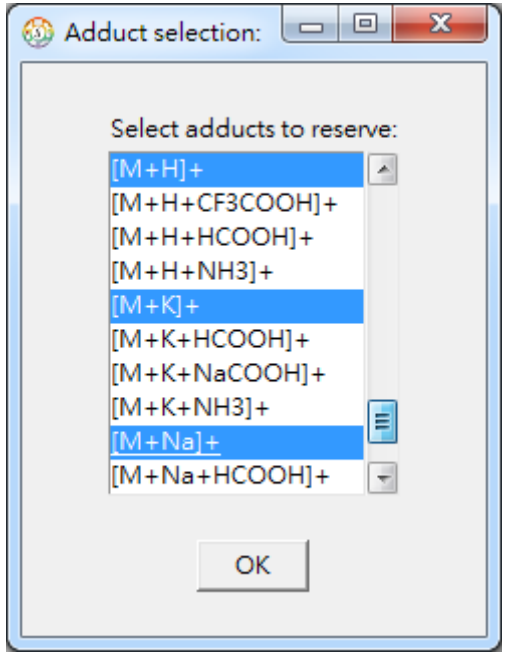
- Re-alignment

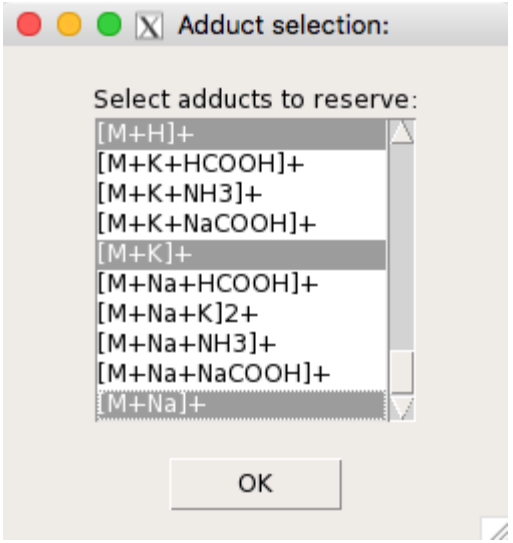
& annotation

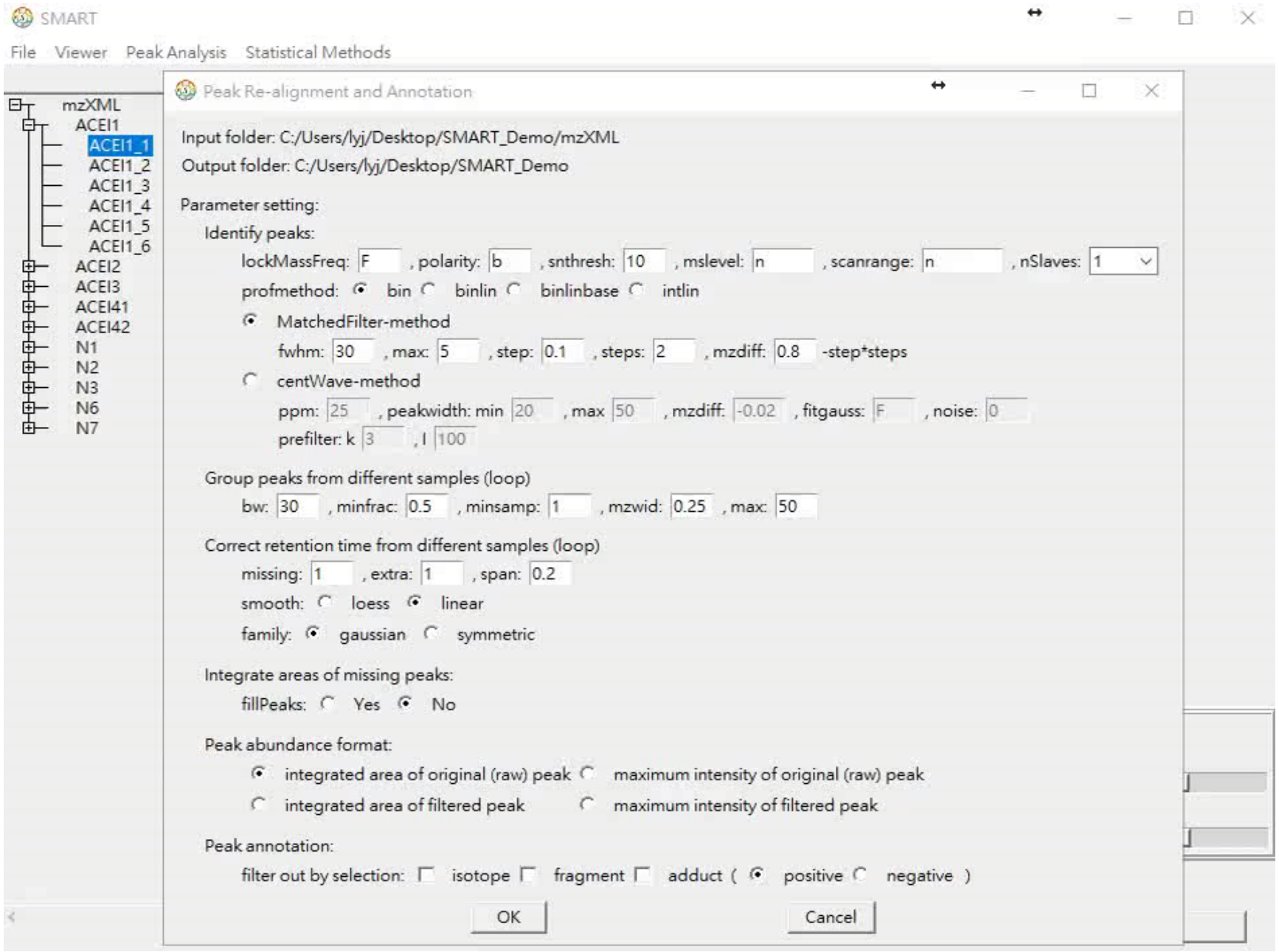


Statistical Metabolomics Analysis – an R Tool

- Re-alignment
  - CAMERA



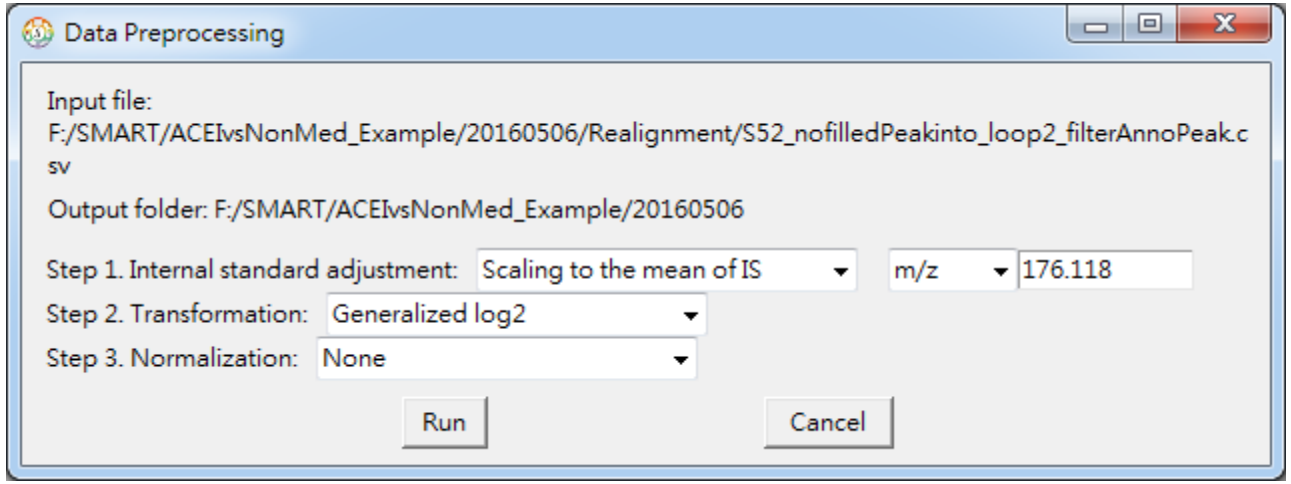




Peak re-alignment

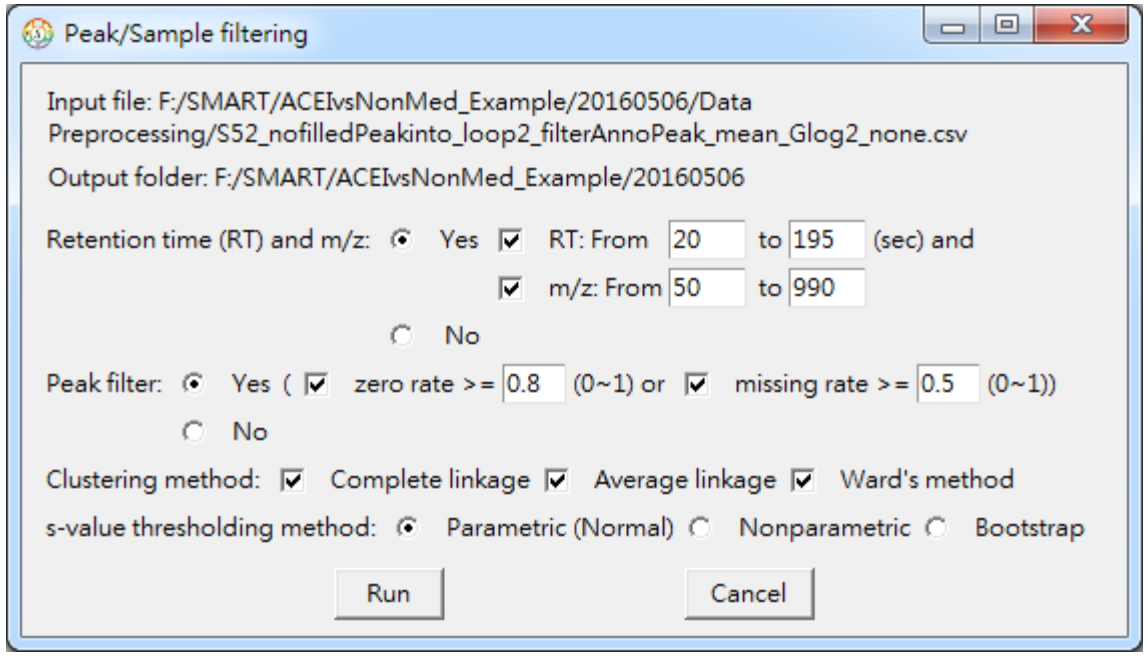
Statistical Metabolomics Analysis – an R Tool

- Data preprocessing (after re-alignment)



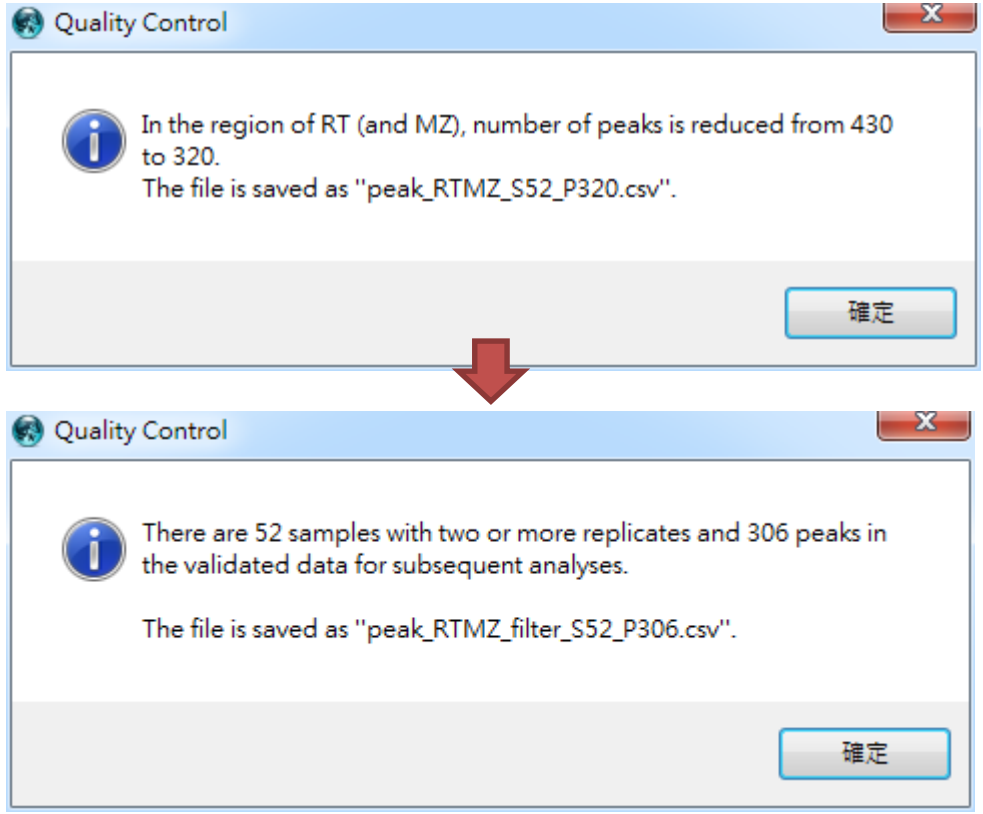
Statistical Metabolomics Analysis – an R Tool

- Quality control (after re-alignment)
  - Peak filtering



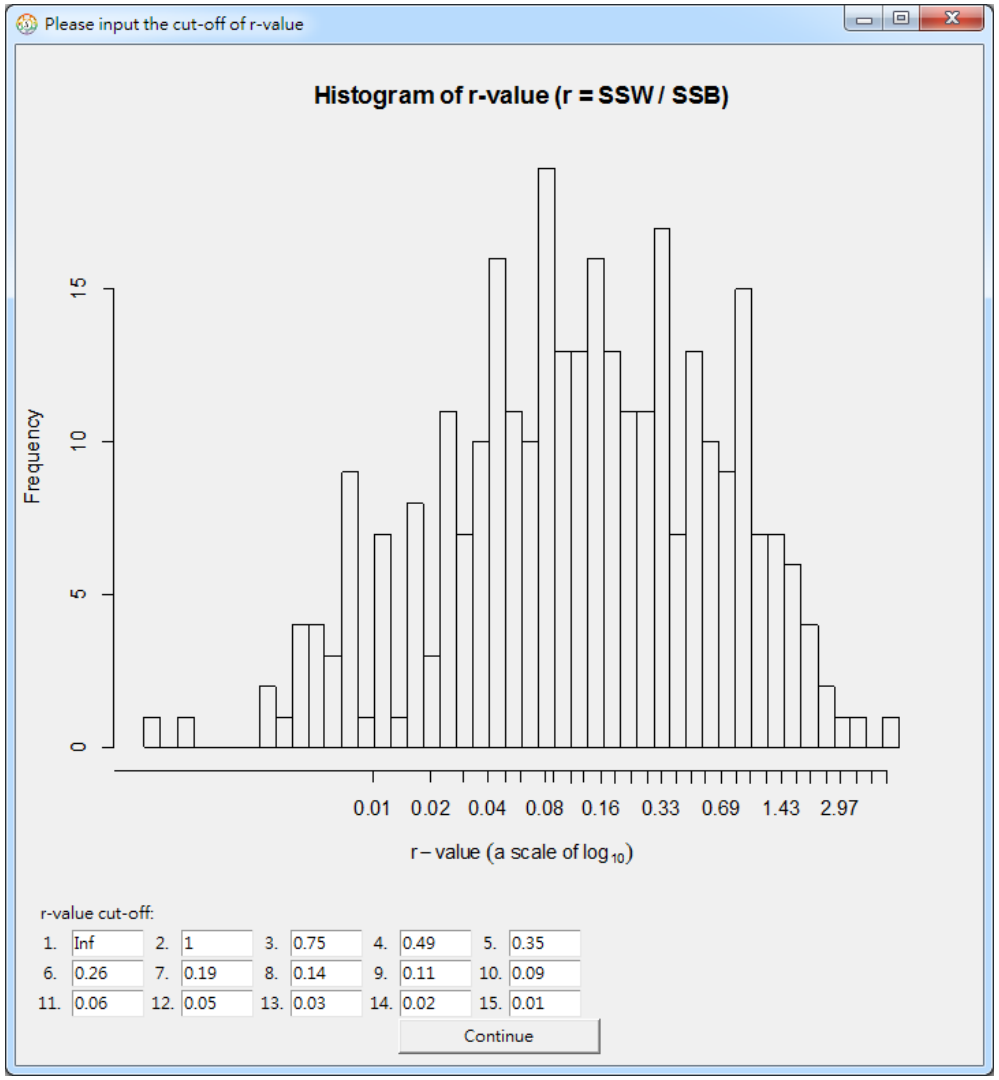
Statistical Metabolomics Analysis – an R Tool

- Quality control (after re-alignment)
  - Peak filtering



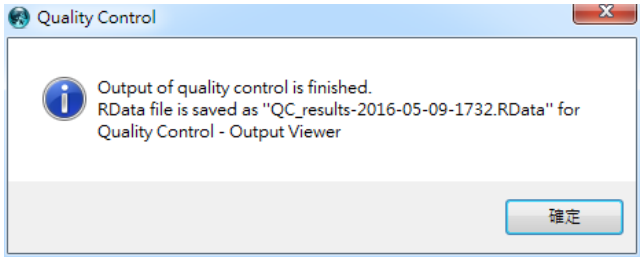
Statistical Metabolomics Analysis – an R Tool

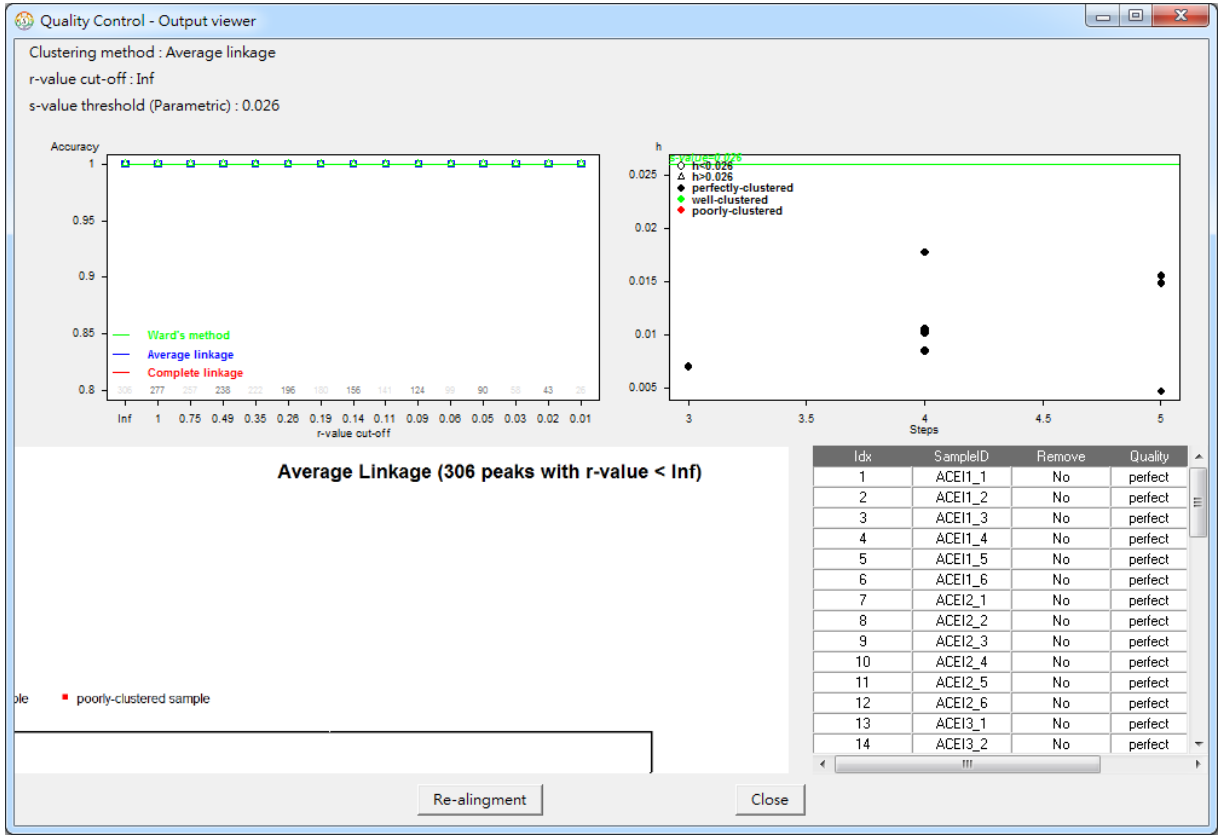
- Quality control (after re-alignment)
  - Peak filtering



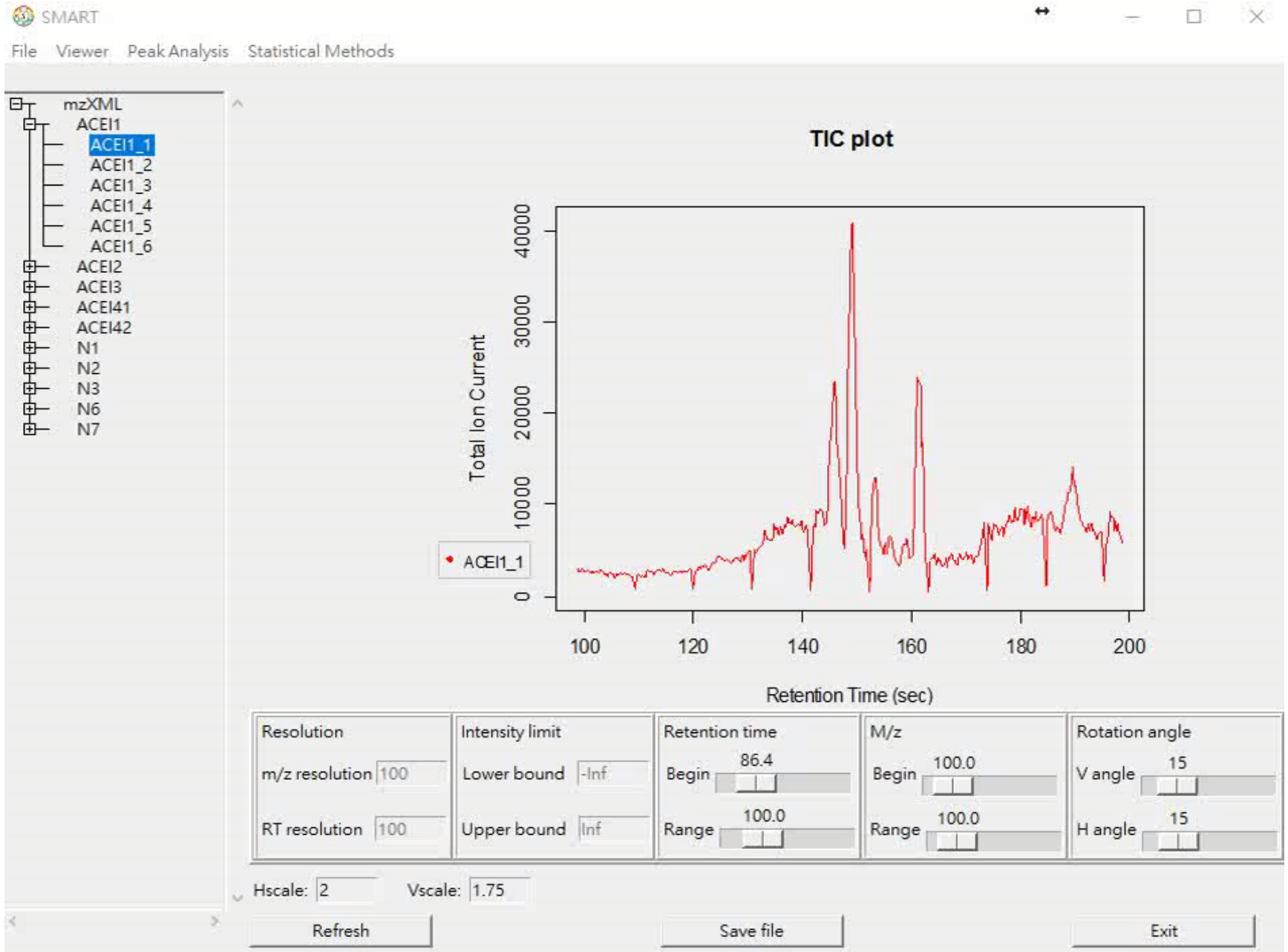
Statistical Metabolomics Analysis – an R Tool

- Quality control (after re-alignment)
  - Peak filtering



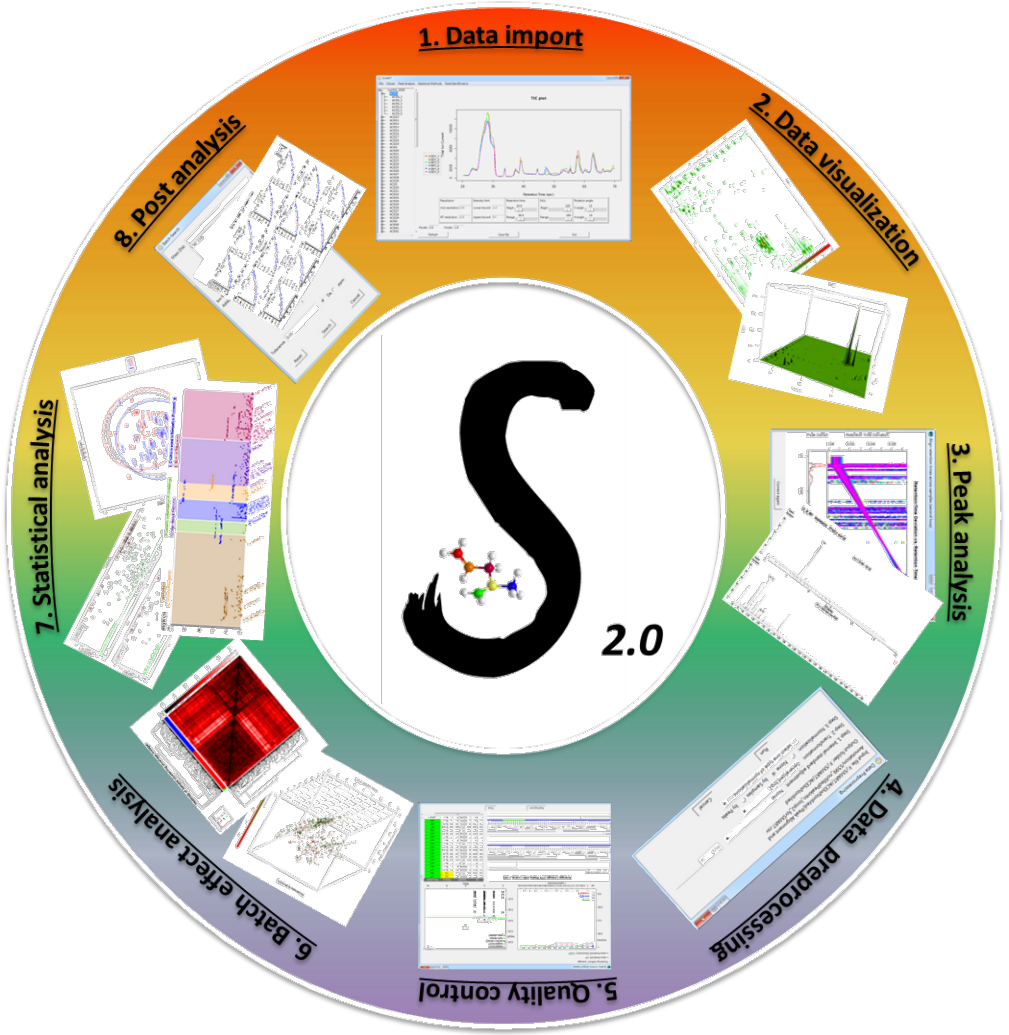


Idx	SampleID	Remove	Quality
1	ACEI1_1	No	perfect
2	ACEI1_2	No	perfect
3	ACEI1_3	No	perfect
4	ACEI1_4	No	perfect
5	ACEI1_5	No	perfect
6	ACEI1_6	No	perfect
7	ACEI2_1	No	perfect
8	ACEI2_2	No	perfect
9	ACEI2_3	No	perfect
10	ACEI2_4	No	perfect
11	ACEI2_5	No	perfect
12	ACEI2_6	No	perfect
13	ACEI3_1	No	perfect
14	ACEI3_2	No	perfect



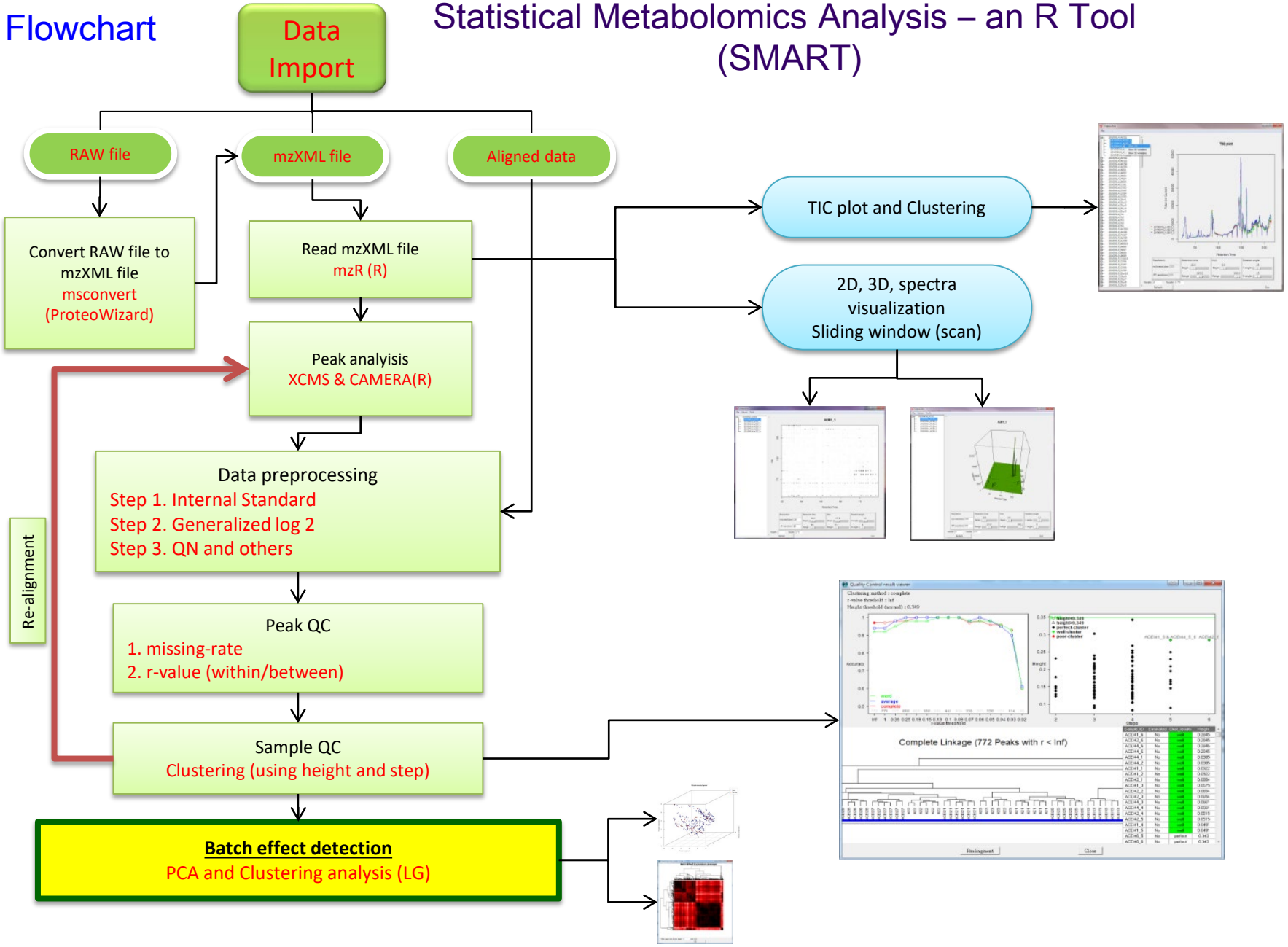
Quality control (2nd)

6. Batch effect detection



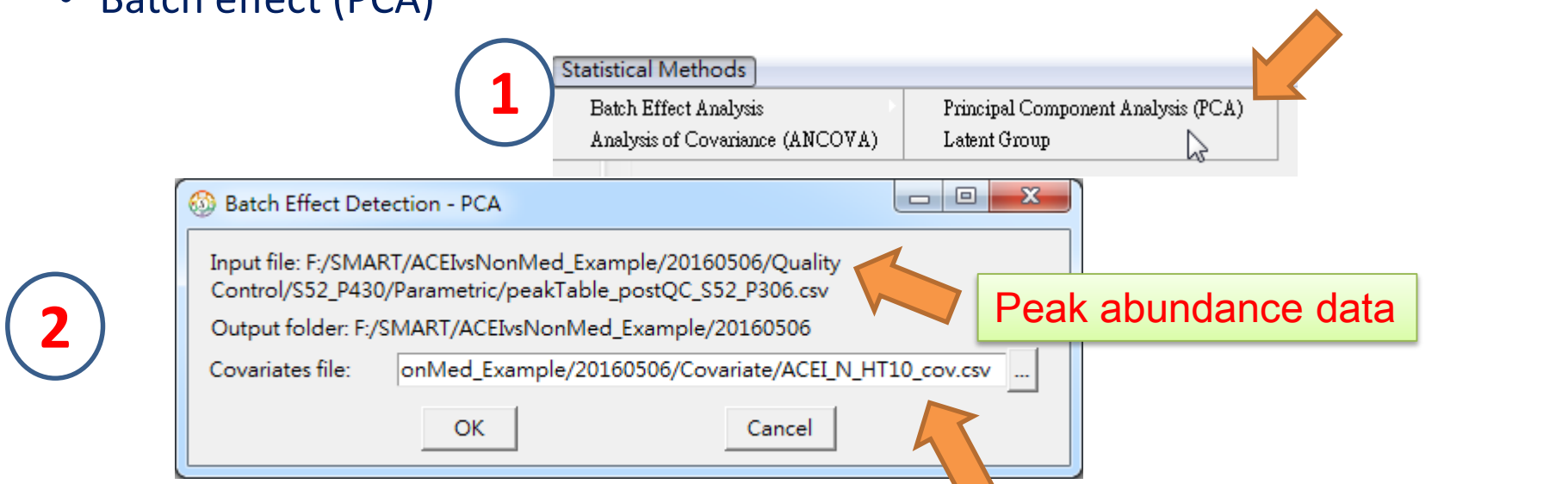
Statistical Metabolomics Analysis – an R Tool (SMART)

Flowchart



Batch effect detection

- Batch effect (PCA)



Peak abundance data (peakTable_postQC.csv)

Peak_Index	mz	Ret_time.sec	ACEI1_1	ACEI1_2
3	60.07563	29.1905	5.02179	5.47943
4	61.01206	39.1727	NA	NA
			6.11894	6.54919
			NA	NA
			NA	NA
8	72.08126	64.0673	NA	NA
9	72.08131	37.8041	7.33726	7.13653
10	77.03941	58.0969	6.26716	6.54659
11	79.05265	58.1402	5.65197	5.21454

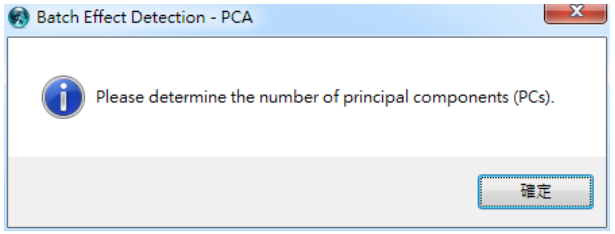
Cov file

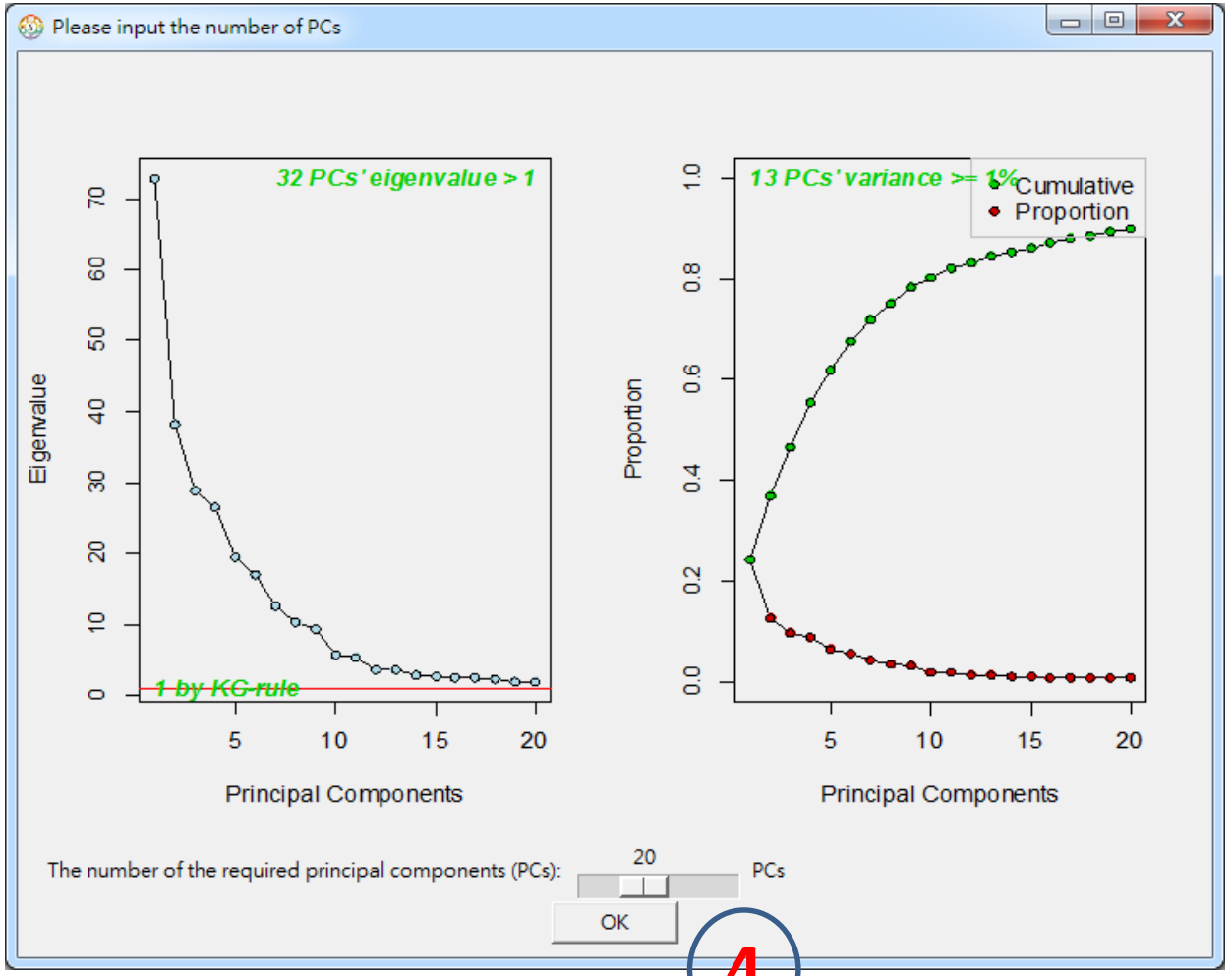
SampleID	DDataset	Dgender	Cage	Cbmi	Ddate
ACEI1	1	1	46	26.08	D0914
ACEI10	1	1	46	26.65	D0915
ACEI11	2	1	46	23.41	D0920
ACEI12	2	1	47	27.32	D0920
ACEI13		1	46	23.76	D0920
ACEI14		1	42	30.81	D1014
ACEI16	2	1	41	23.29	D0924
ACEI17	2	1	23	25.45	D0924
ACEI18	2	1	24	29	D0924
ACEI19	2	2	50	26.08	D0924

Batch effect detection

- Batch effect (PCA)

3





4

Batch effect detection

- Batch effect (PCA)
  - Output files

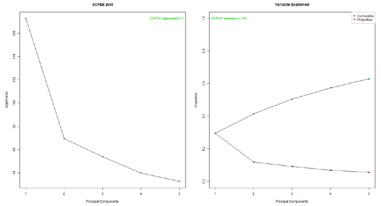
outpath\Batch Effect Detection\Principal Component Analysis (5)

SampleID	PC_1	PC_2	PC_3
ACEI1_1	-14.42	-8.34195	-4.72467
ACEI1_2	-15.9665	-8.26392	-1.82812
ACEI1_3	-10.3364	-12.6445	-11.1373
ACEI1_4	-9.78977	-11.5955	-9.99943
ACEI1_5	-8.3899	-10.1633	-5.99414
ACEI1_6	-10.9973	-4.66903	-6.23839
ACEI2_1	17.64336	-6.43721	6.890438

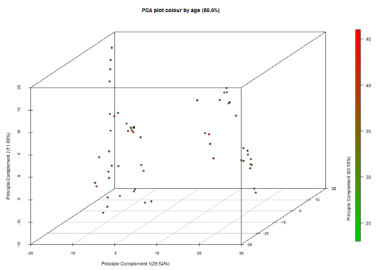
PCA_scores_5.csv

PC_index	PC_eigenvalue	PC_varprop	PC_varpropcum
1	172.8498312	0.295427713	0.295427713
2	69.48399819	0.118759148	0.414186861
3	53.71048208	0.091799713	0.505986574
4	39.69078614	0.067837834	0.573824408
5	32.63368001	0.055776123	0.629600531

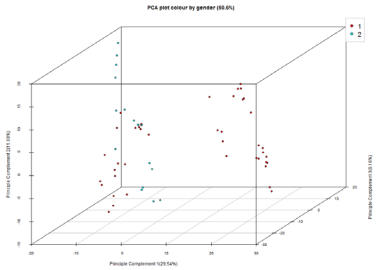
PCA_variance_5.csv



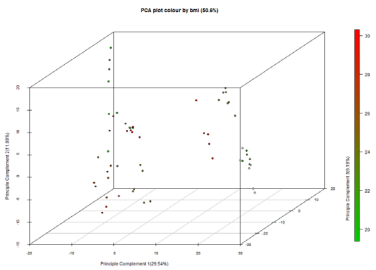
PCA_screeplot_5.png



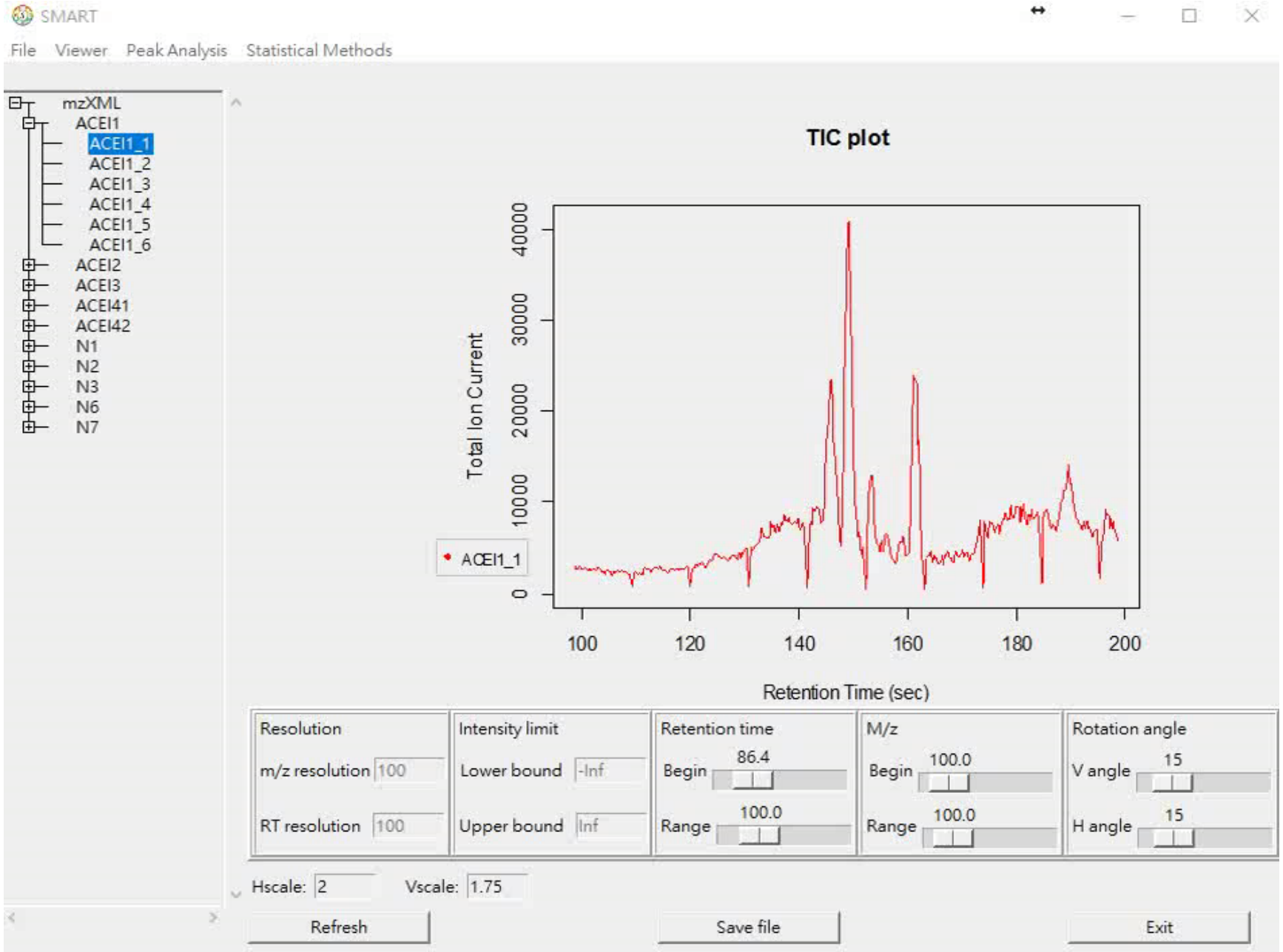
PCAplot_colourby_age (50.6%).png



PCAplot_colourby_gender (50.6%).png



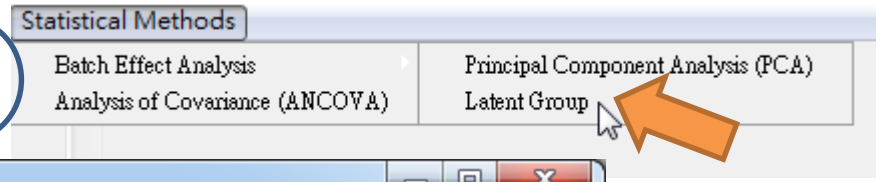
PCAplot_colourby_BMI (50.6%).png

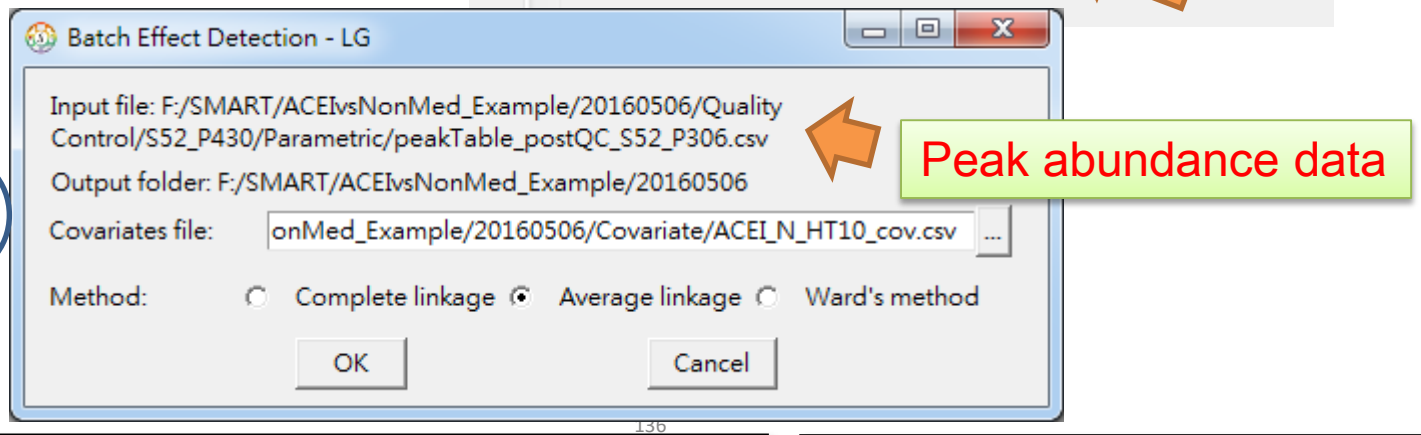


Batch effect (PCA)

Batch effect detection

- Batch effect (latent group)
  - Clustering analysis





Peak abundance data (peakTable_postQC.csv)

Peak_Index	Mass	Ret_time(sec)	N7_1	N7_6
3	60.0756	29.1905	4.9484	NA
4	61.0121	39.1727	NA	NA
5	68.9834	23.9602	7.40761	7.19992
			NA	NA
			NA	NA
			NA	NA
9	72.0813	37.8041	6.70913	5.64685
10	77.0394	58.0969	6.34075	5.80933
11	79.0526	58.1402	NA	4.92574

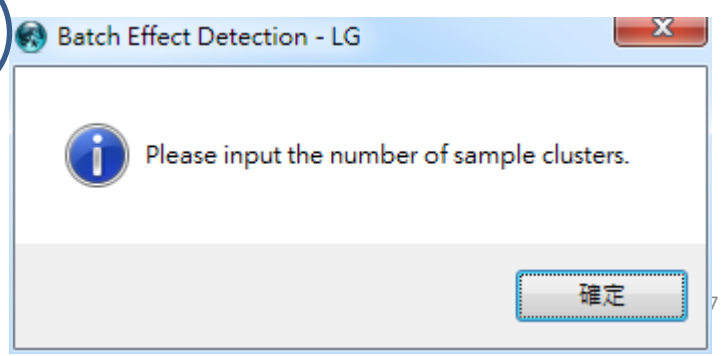
Cov file

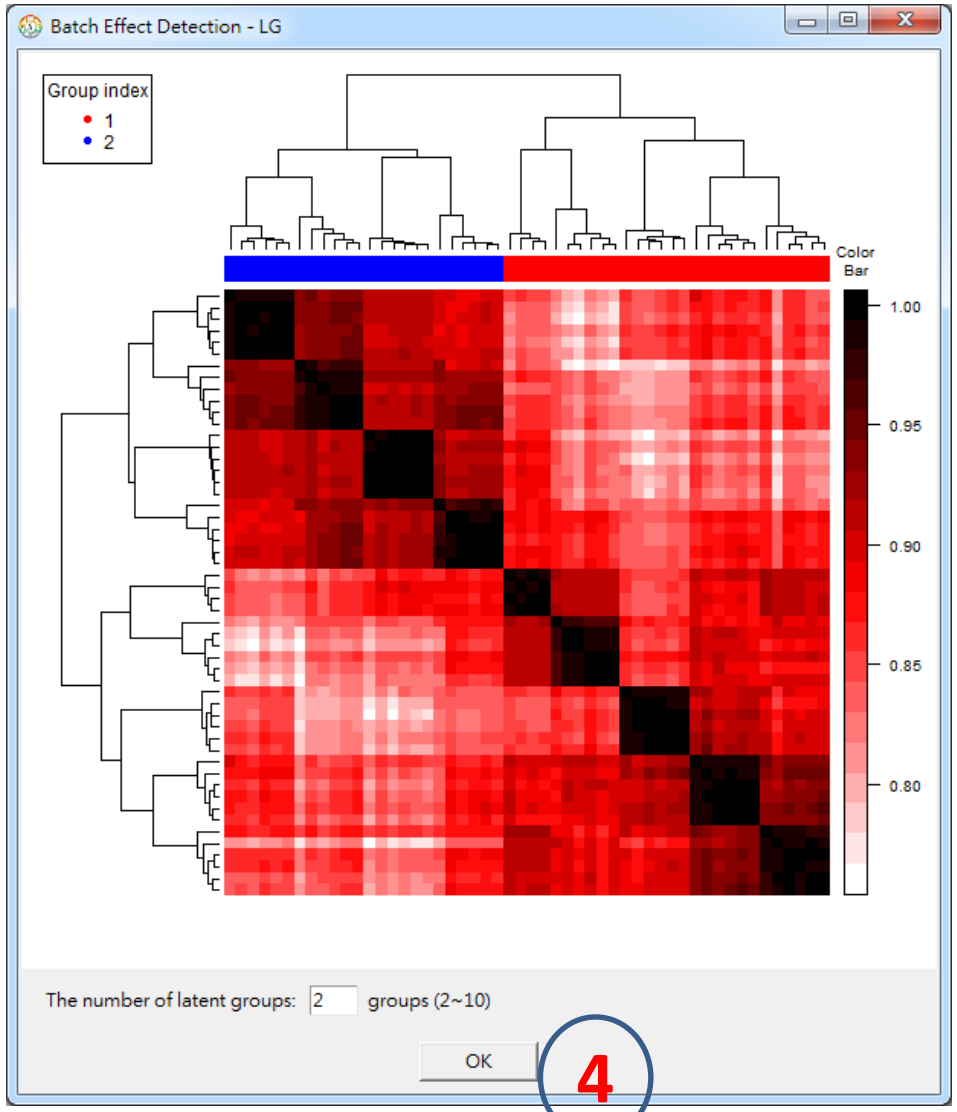
SampleID	DDataset	Dgender	Cage	Cbmi	Ddate
ACEI1	1	1	46	26.08	D0914
ACEI10	1	1	46	26.65	D0915
ACEI11	2	1	46	23.41	D0920
ACEI12		1	47	27.32	D0920
ACEI13		1	46	23.76	D0920
ACEI14	5	1	42	30.81	D1014
ACEI16	2	1	41	23.29	D0924
ACEI17	2	1	23	25.45	D0924
ACEI18	2	1	24	29	D0924
ACEI19	2	2	50	26.08	D0924

Batch effect detection

- Batch effect (latent group)
  - Clustering analysis

3





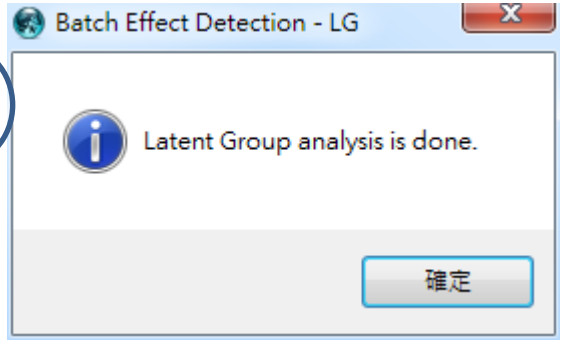
4

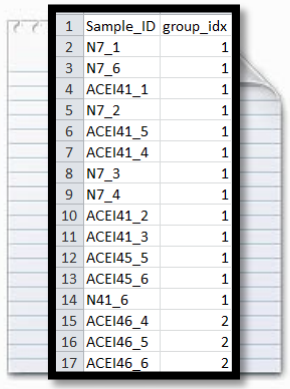
Batch effect detection

- Batch effect (latent group)
  - Clustering analysis (output files)

outpath\Batch Effect Analysis\Latent Group (2)

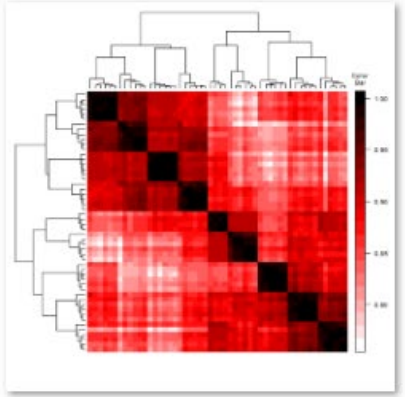
5



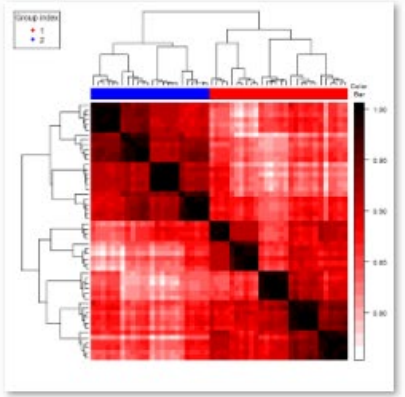


	Sample_ID	group_idx
2	N7_1	1
3	N7_6	1
4	ACEI41_1	1
5	N7_2	1
6	ACEI41_5	1
7	ACEI41_4	1
8	N7_3	1
9	N7_4	1
10	ACEI41_2	1
11	ACEI41_3	1
12	ACEI45_5	1
13	ACEI45_6	1
14	N41_6	1
15	ACEI46_4	2
16	ACEI46_5	2
17	ACEI46_6	2

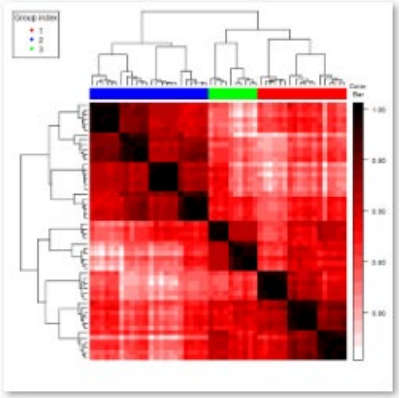
latent_group_Average_2.csv



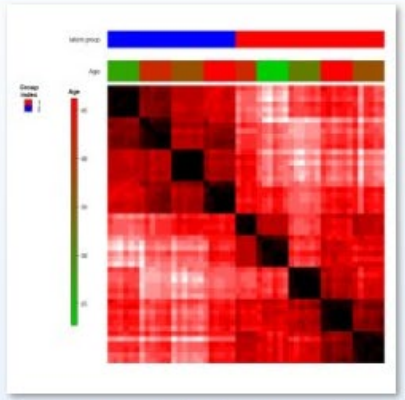
latent_group_Average.png



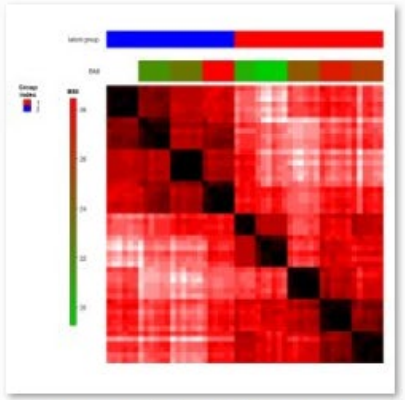
latent_group_Average_2.png



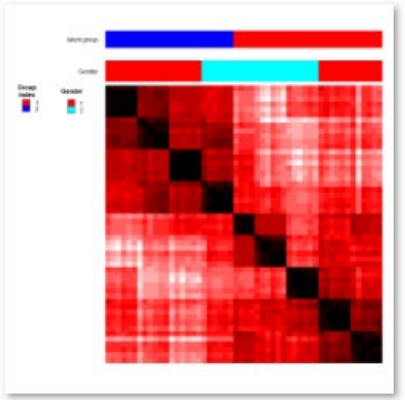
latent_group_Average_3.png



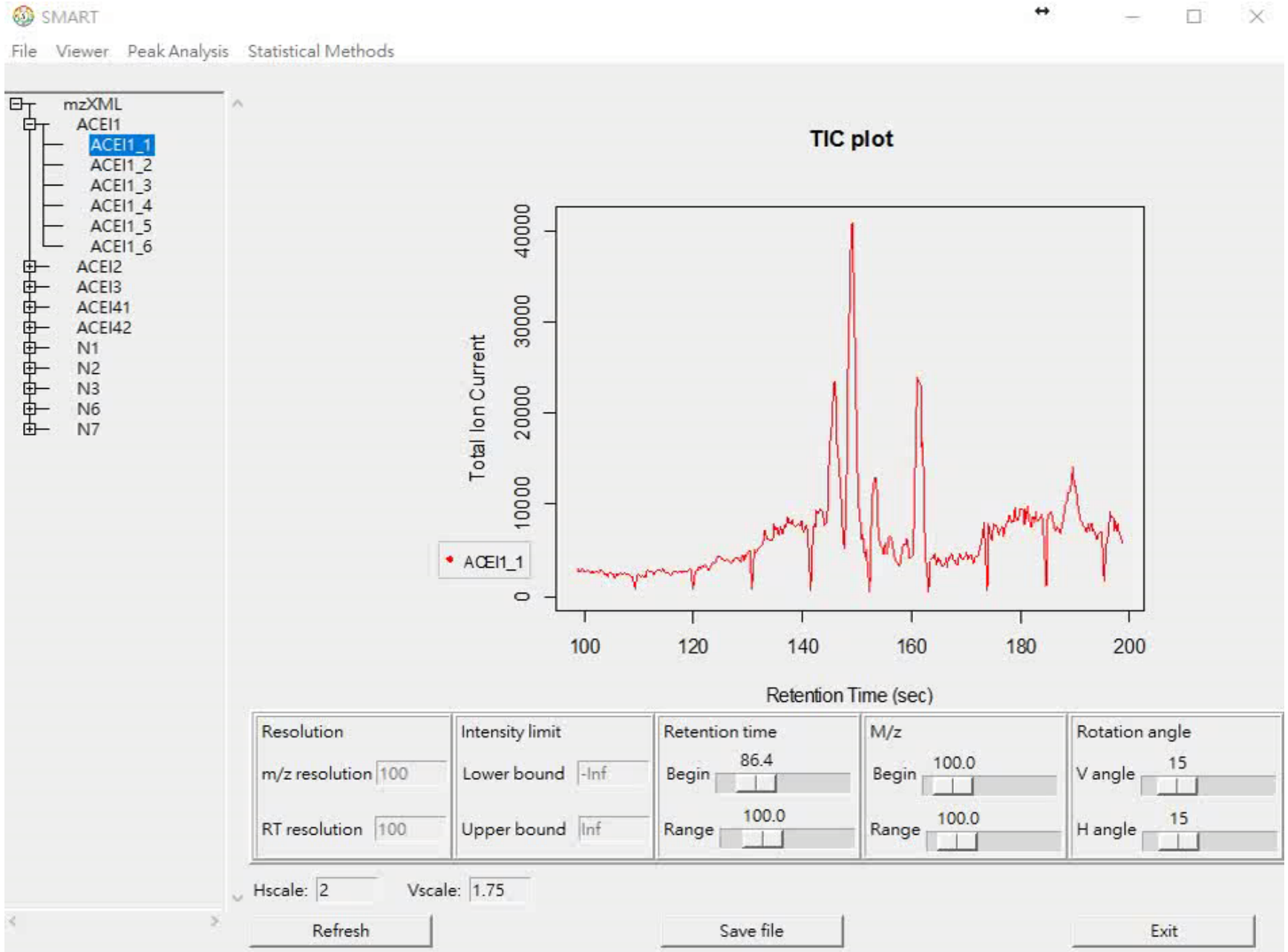
latentgroup_colorby_Age_Average_2.png



latentgroup_colorby_BMI_Average_2.png

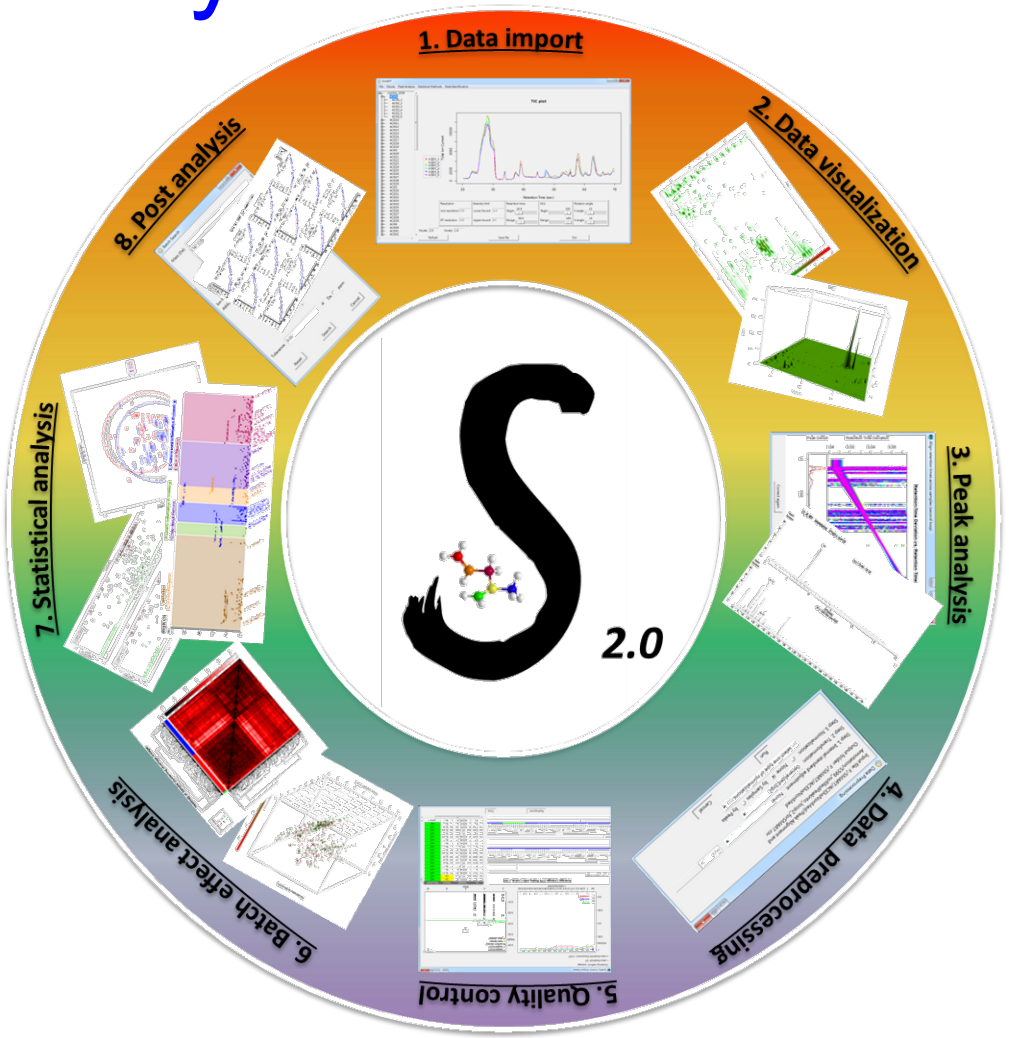


latentgroup_colorby_Gender_Average_2.png



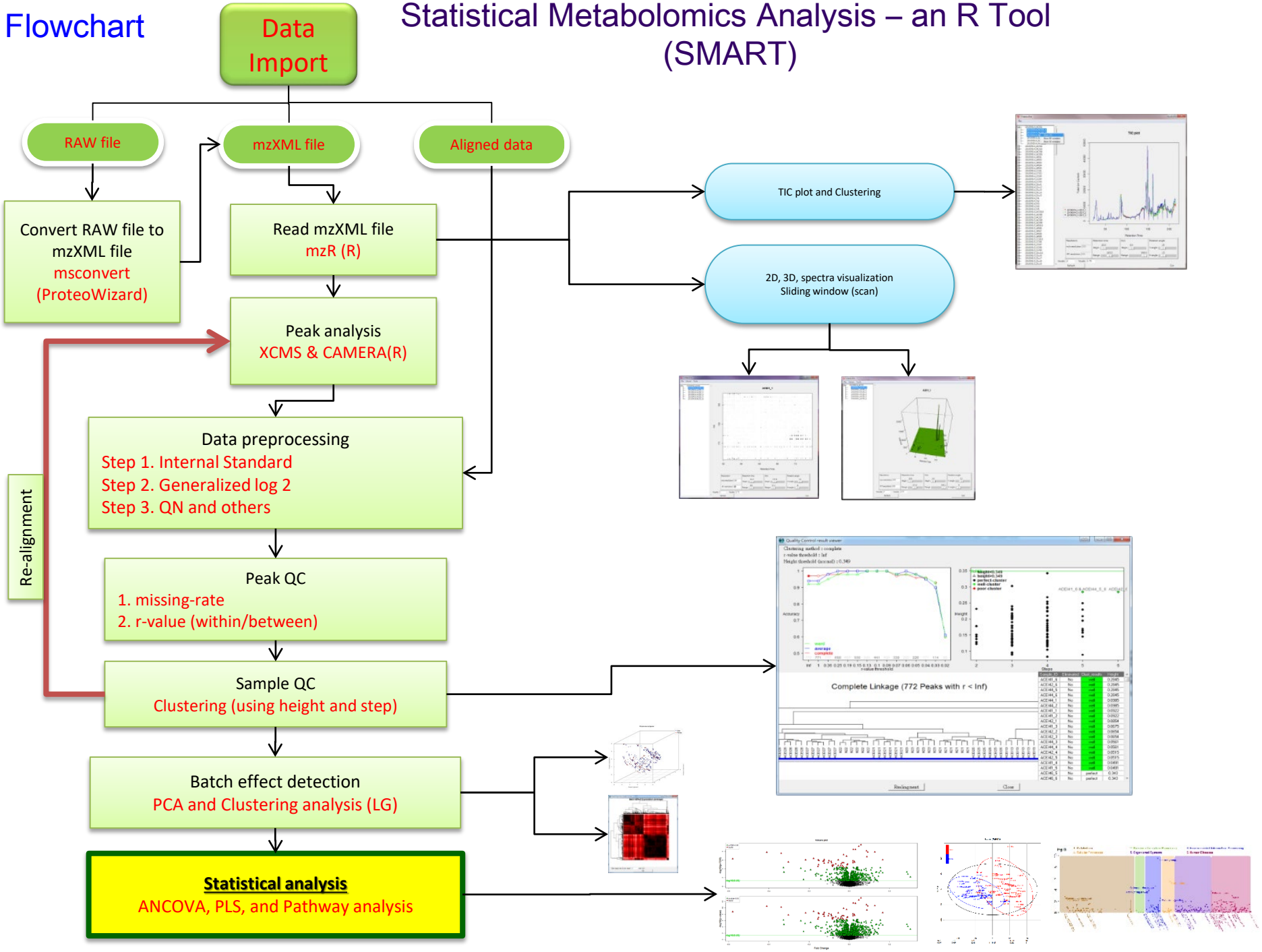
Batch effect (latent group)

7. Statistical analysis



Statistical Metabolomics Analysis – an R Tool (SMART)

Flowchart



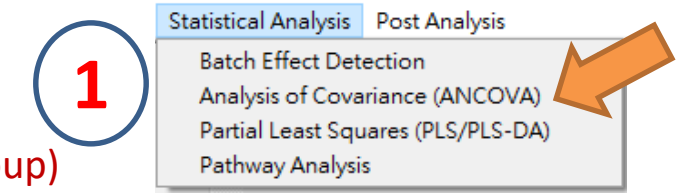
ANCOVA

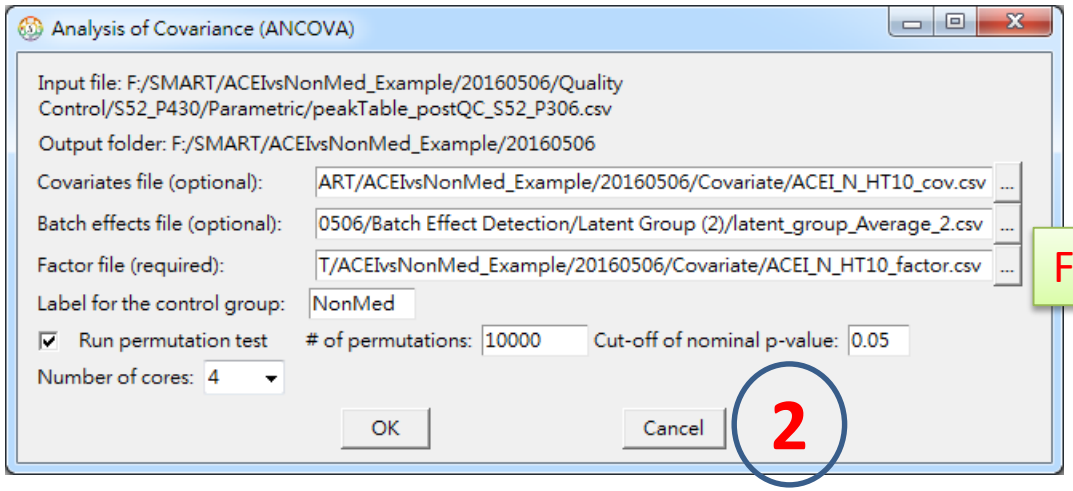
STATISTICAL ANALYSIS (1)

Statistical analysis (ANCOVA)

Latent group

- Association analysis
  - ANCOVA
  - Batch effect adjustment (Latent group)





Factor (variable of interest) file

Cov file

SampleID	DDataset	Dgender	Cage	Cbmi	Ddate
ACEI1	1	1	46	26.08	D0914
ACEI10	1	1	46	26.65	D0915
ACEI11	2	1	46	23.41	D0920
ACEI12	2	1	47	27.32	D0920
ACEI13	2	1	46	23.76	D0920
ACEI14	5	1	42	30.81	D1014
ACEI16	2	1	41	23.29	D0924
ACEI17	2	1	23	25.45	D0924
ACEI18	2	1	24	29	D0924
ACEI19	2	2	50	26.08	D0924

Batch effect file (latent group)

	Sample_ID	group_idx
2	N7_1	1
3	N7_6	1
4	ACEI41_1	1
5	N7_2	1
6	ACEI41_5	1
7	ACEI41_4	1
8	N7_3	1
9	N7_4	1
10	ACEI41_2	1
11	ACEI41_3	1

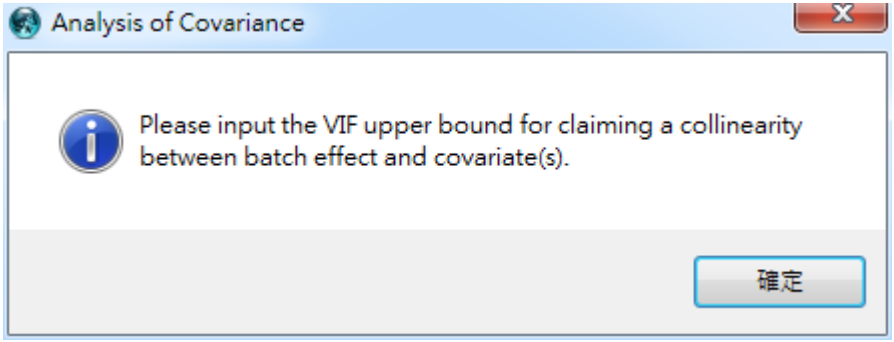
Factor file

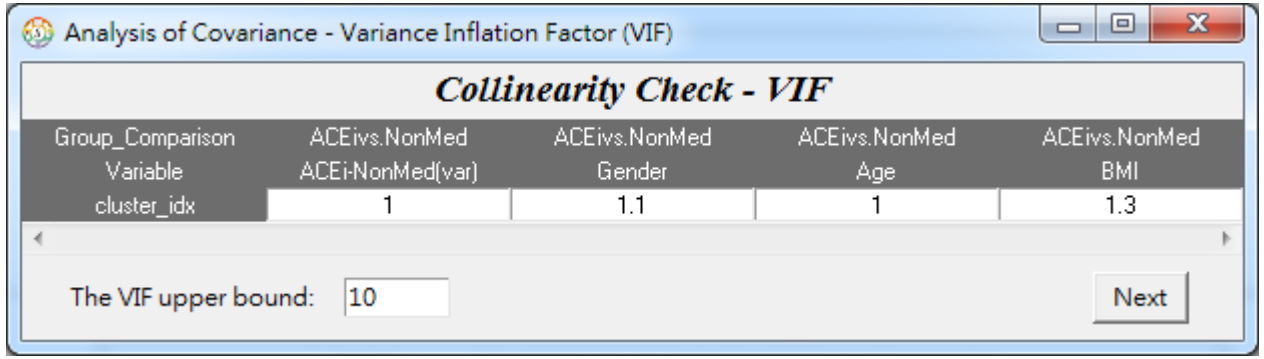
SampleID	Group
ACEI1	ACEI
ACEI2	ACEI
ACEI3	ACEI
ACEI4	ACEI
ACEI5	ACEI
ACEI6	ACEI
ACEI7	ACEI
ACEI8	ACEI
ACEI9	ACEI
ACEI10	ACEI

Statistical analysis (ANCOVA)

Latent group

- Association analysis
  - ANCOVA
    - VIF

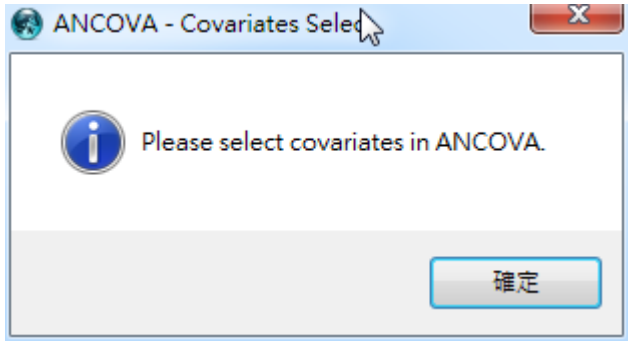


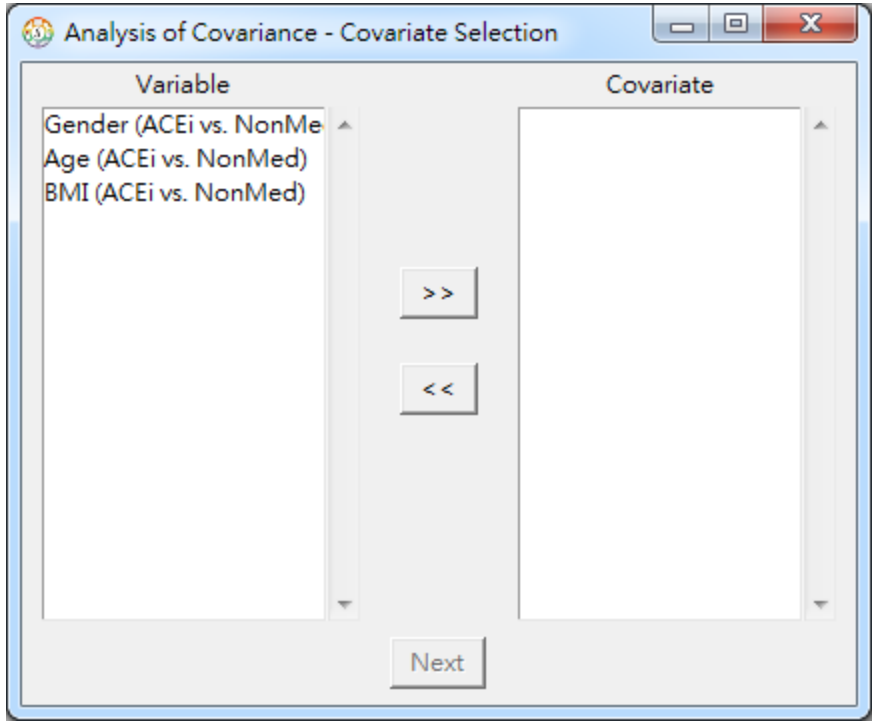


Statistical analysis (ANCOVA)

Latent group

- Association analysis
  - ANCOVA
    - Covariate selection

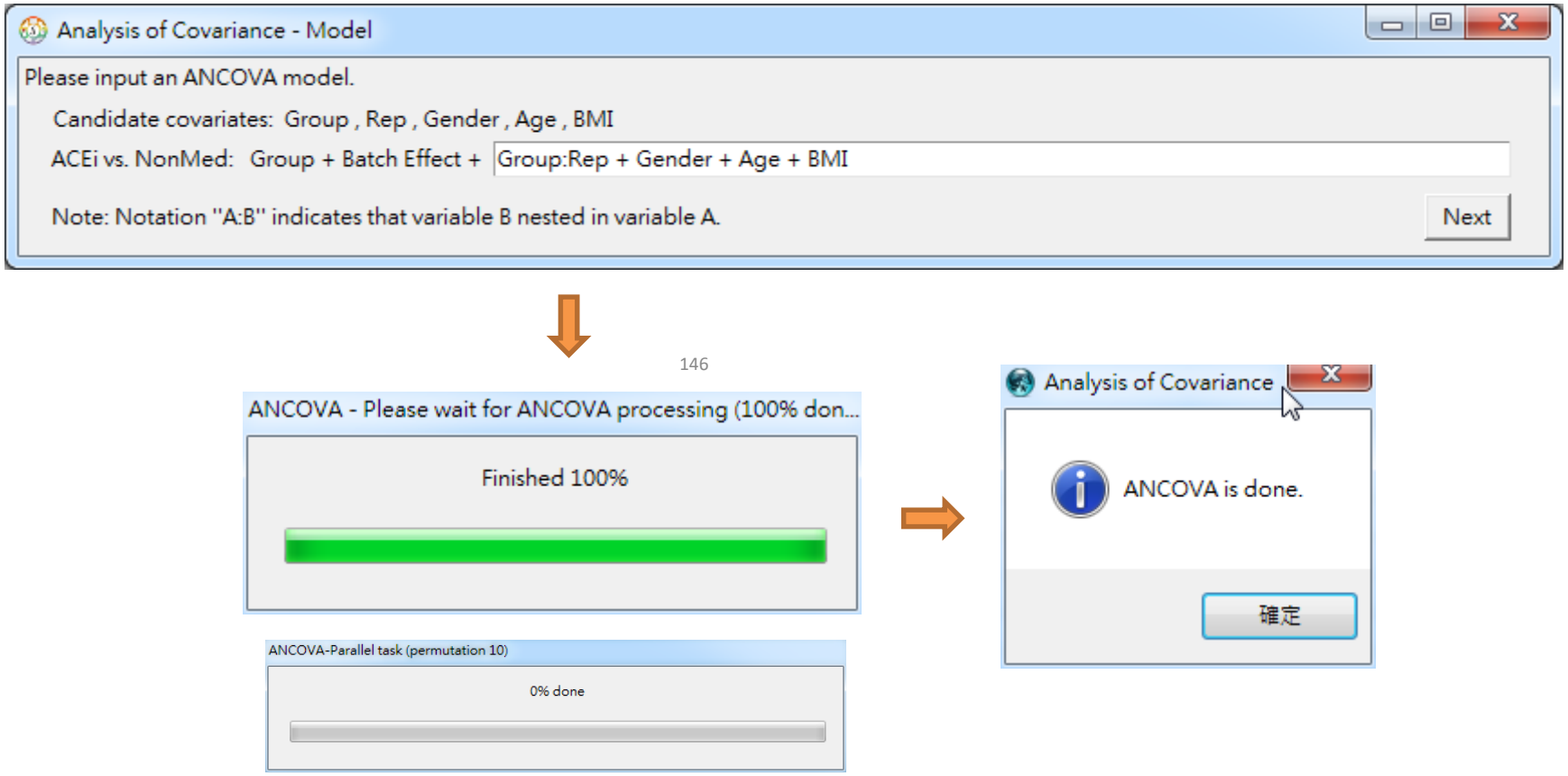




Statistical analysis (ANCOVA)

Latent group

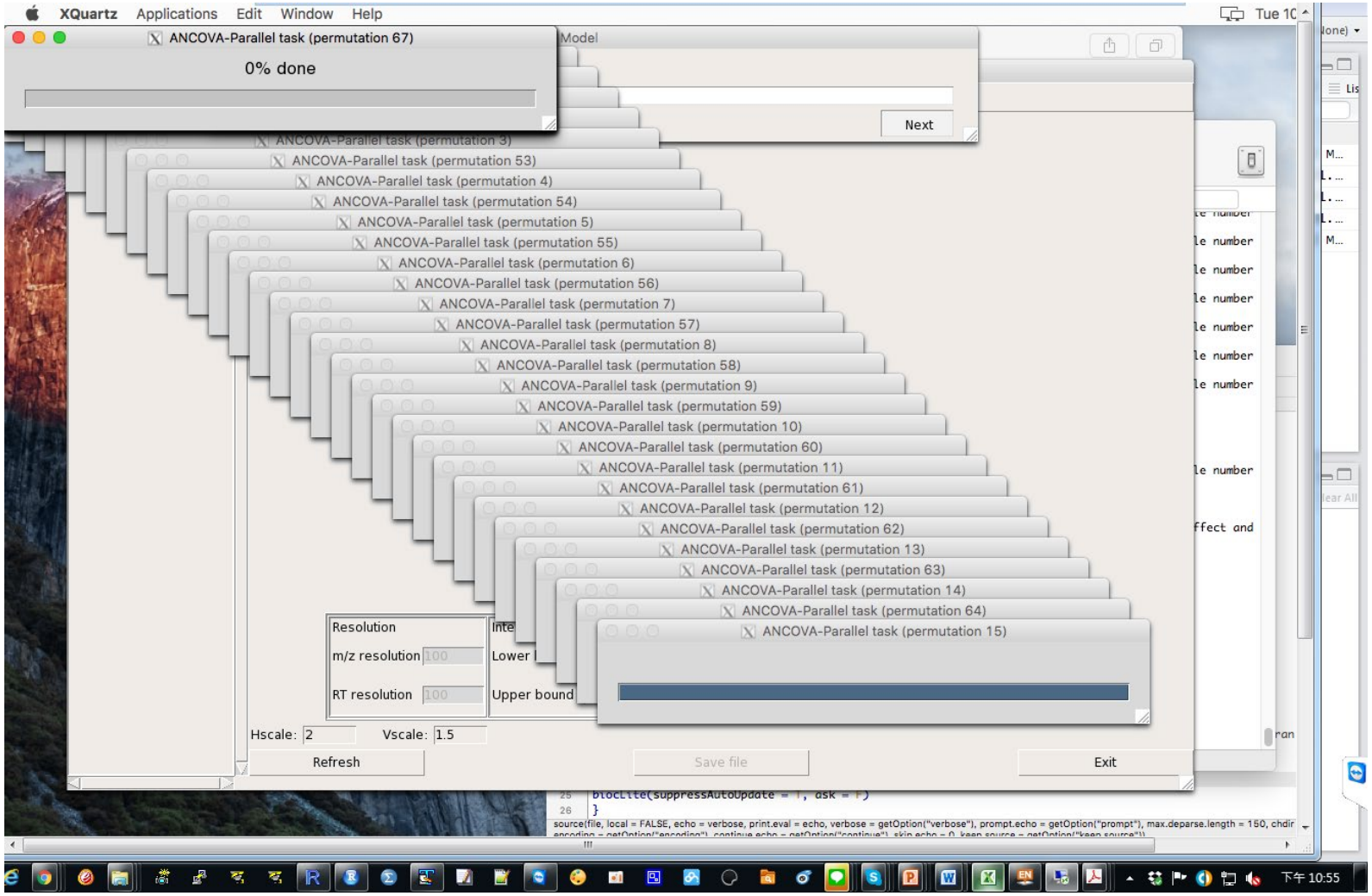
- Association analysis
  - ANCOVA
    - Model setting : Group:Rep + Gender + Age + BMI



Statistical analysis (ANCOVA)

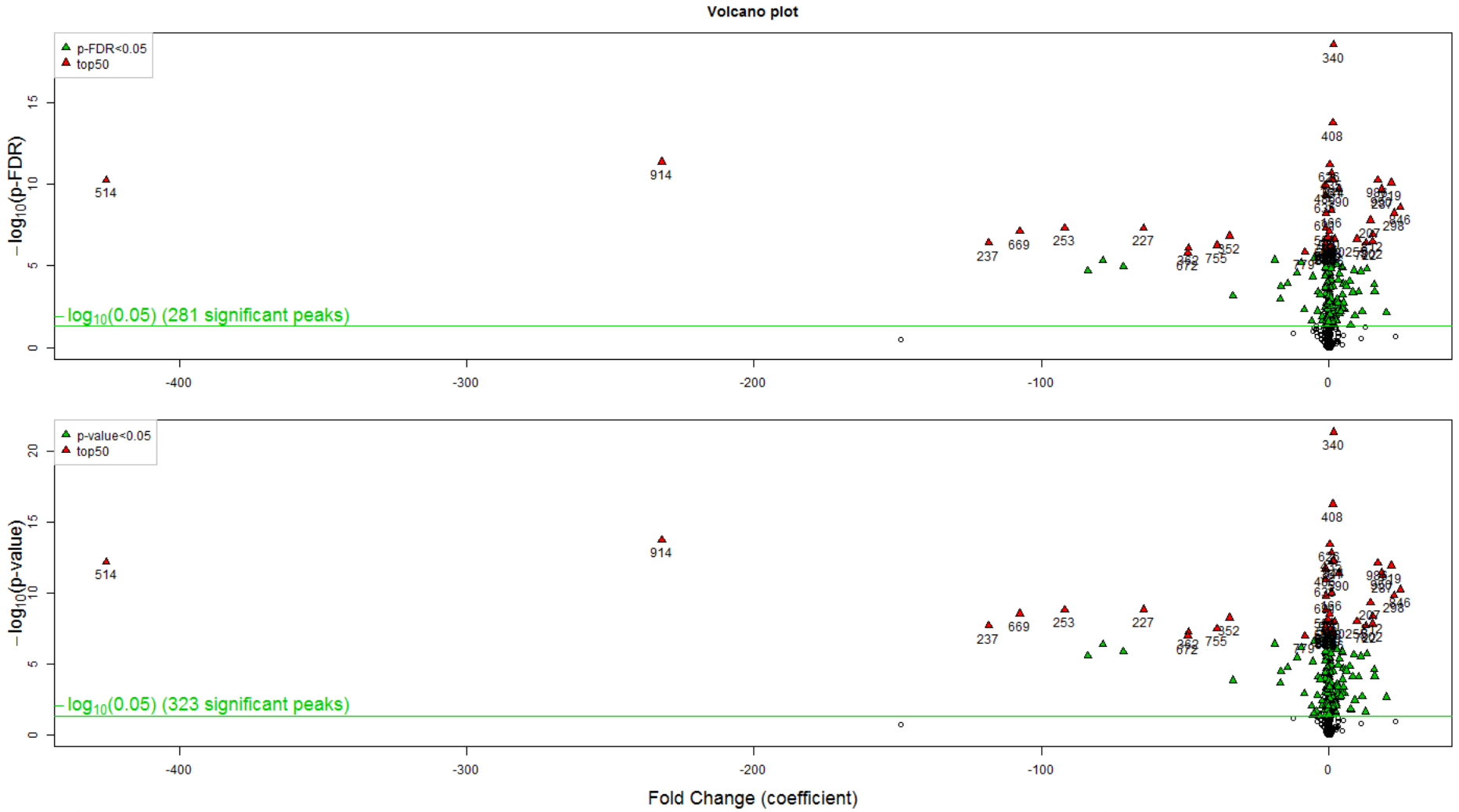
Latent group

- Association analysis
  - ANCOVA
    - Permutation



Statistical analysis (ANCOVA)

- Association analysis
  - ANCOVA output (Volcano plot)



outpath\ANCOVA\Volcanoplot_ACEi vs. NonMed.png

Statistical analysis (ANCOVA)

- Association analysis
  - ANCOVA output (numerical results)

outpath\ANCOVA\

VIF_threshold_10.csv

Group_Comparison	ACEivs.NonMed	ACEivs.NonMed	ACEivs.NonMed	ACEivs.NonMed
Variable	ACEi-NonMed(var)	gender	age	bmi
cluster_idx	1.1	1.1	1	1

covariate_select.csv

	ACEi
Dataset	0
gender	1
age	1
bmi	1
date	0

coef_ACEi vs. NonMed.csv

Peak_Index	mz	Ret_time.sec	ACEi_NonMed
4	60.07346	29.3075	0.193368544
5	68.97672	24.1263	-0.275392306
6	70.05911	32.3293	-13.3419297
7	72.07476	37.8294	-0.220116074
8	72.07494	64.0451	-6.089761406

p-value_ACEi vs. NonMed_details.csv

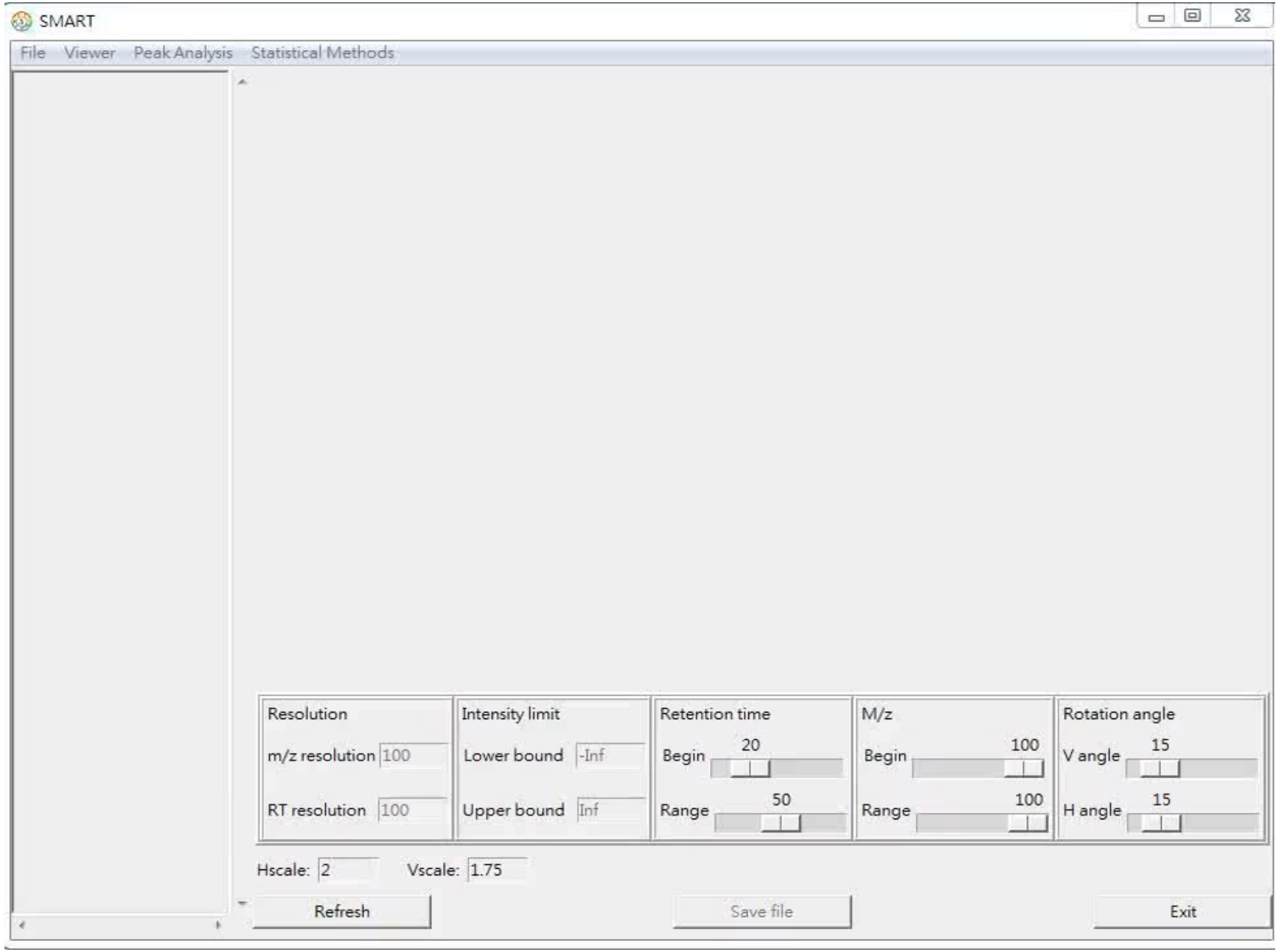
Peak_Index	mz	Ret_time.sec	(Intercept)	ACEi_NonMed	Gender	Age	BMI	cluster_idx	ACEi_NonMed:Rep
6	60.07116129	29.2696	1.07E-11	0.964754576	0.004593074	0.10313761	0.102492534	0.000784103	0.894761782
8	69.06404612	52.8168	0.000887385	0.001643264	0.002529177	0.007070659	0.000141511	NA	0.374987713
9	70.05927741	32.3226	0.01042616	0.005719337	0.013706893	0.000968562	NA	NA	0.024744527
20	84.0381544	48.7662	1.65E-05	0.44412925	0.048497137	0.482746084	0.953953795	4.00E-05	0.979548774
24	84.95377894	24.1371	4.03E-33	0.001005549	3.23E-11	4.07E-12	5.22E-19	4.04E-28	0.99475372

pFDR_ACEi vs. NonMed.csv

Peak_Index	mz	Ret_time.sec	ACEi_NonMed(pfdr)
4	60.07346317	29.3075	0.881473536
5	68.97671978	24.1263	0.14360528
6	70.05911424	32.3293	0.77584518
7	72.0747612	37.8294	0.020267411
8	72.07493563	64.0451	0.745393188

Empirical p-value_ACEi vs. NonMed

Peak_Index	mz	Ret_time.sec	ACEi_NonMed(epv)	ACEi_NonMed(epFDR)
8	69.06404612	52.8168	0.00990099	0.022442244
9	70.05927741	32.3226	0.02970297	0.059020188
24	84.95377894	24.1371	0.00990099	0.022442244
31	88.94712657	24.1306	0.00990099	0.022442244
51	104.1014407	29.0027	0.00990099	0.022442244



ANCOVA(latent group)

PLS/PLS-DA

STATISTICAL ANALYSIS (2)

Statistical analysis (PLS/PLS-DA)

- Breast cancer data

Gene expression data

Peak_index	MZ	RT	GeneS	LHC3265	LHC3630	LHC4662	LHC5838	LHC6065
1	1	1	OR4F17	3.7453	5.3542	3.9132	3.9697	4.1630
2	2	2	14-Sep	8.1687	8.9836	8.6854	8.8927	8.9830
3	3	3	OR4F16	5.7183	7.7406	7.7985	7.0605	7.0725
4	4	4	GPAM	10.8207	8.1991	10.6819	8.3634	10.9824
5	5	5	LOC100287934	7.9763	8.1870	7.9481	8.3095	8.1858
6	6	6	FAM87A	6.0228	5.7201	5.7764	5.7067	5.9063
7	7	7	LOC643837	7.7452	6.6167	8.9223	7.5027	8.1036
8	8	8	SAMD11	7.0537	6.6805	6.9922	6.6381	6.8894
9	9	9	KLHL17	7.6677	7.2229	7.4896	7.0800	7.3386
10	10	10	PLEKHN1	7.5829	7.2255	7.6950	7.3089	7.3336

Cov file

SampleID	CAGE	DRACE/ETHNICITY
LHC3265	51	W
LHC3630	39	B
LHC3660	42	W
LHC4662	91	B
LHC5838	42	W
LHC6065	39	B
LHC6075	81	B

Factor file

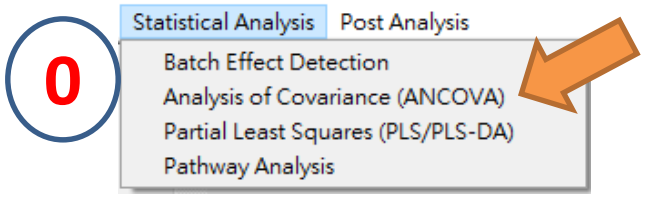
SampleID	DGroup
LHC3265	NORMAL
LHC3630	NORMAL
LHC3660	NORMAL
LHC4662	NORMAL
LHC3264	TUMOR
LHC3631	TUMOR
LHC3661	TUMOR
LHC4333	TUMOR

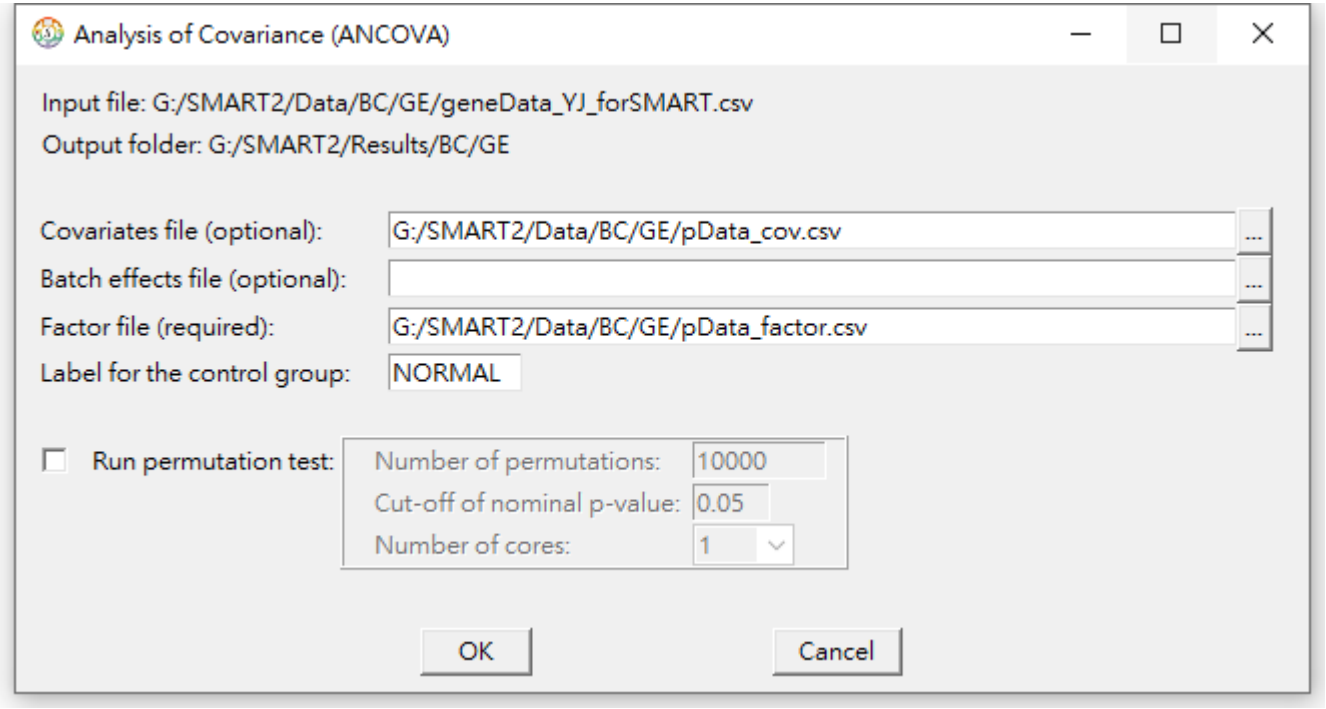
Metabolomics data

Peak_index	MZ	RT	Metabolite	LHC3265	LHC3630	LHC3660	LHC4662	LHC5838
1	1	1	1-arachidonoylglycerophosphoethanolamine*	-1.4956	-0.8435	-0.4623	-1.7410	0.3248
2	2	2	1-arachidonoylglycerophosphoinositol*	-1.2812	-1.2812	-0.7580	-1.2812	0.1697
3	3	3	1-linoleoylglycerophosphocholine	-2.7585	-2.7585	-0.1364	-2.7585	-2.7585
4	4	4	1-methylimidazoleacetate	-0.8876	-0.8876	-0.8876	-0.8876	-0.8876
5	5	5	1-methylnicotinamide	-1.2717	-2.1301	-0.2701	-2.5847	0.7415
6	6	6	1-myristoylglycerophosphocholine	-1.9470	-1.9470	-1.9470	-1.9470	-1.9470
7	7	7	1-oleoylglycerol (1-monoolein)	-4.8191	-4.8191	-4.8191	-4.8191	-4.8191
8	8	8	1-oleoylglycerophosphocholine	-3.3124	-3.3124	-0.4852	-3.3124	-3.3124
9	9	9	1-oleoylglycerophosphoethanolamine	-3.3418	-3.3418	-3.3418	-3.3418	-3.3418
10	10	10	1-palmitoylglycerol (1-monopalmitin)	-3.2395	-3.2395	-1.6936	-3.2395	-3.2395

Statistical analysis (PLS/PLS-DA)

- PLS-DA
  - ANCOVA



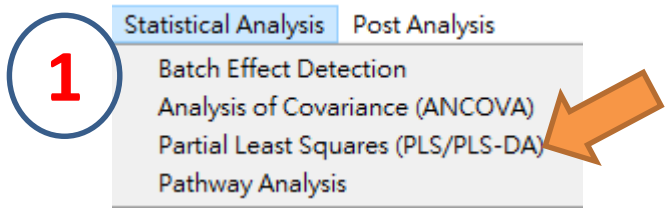


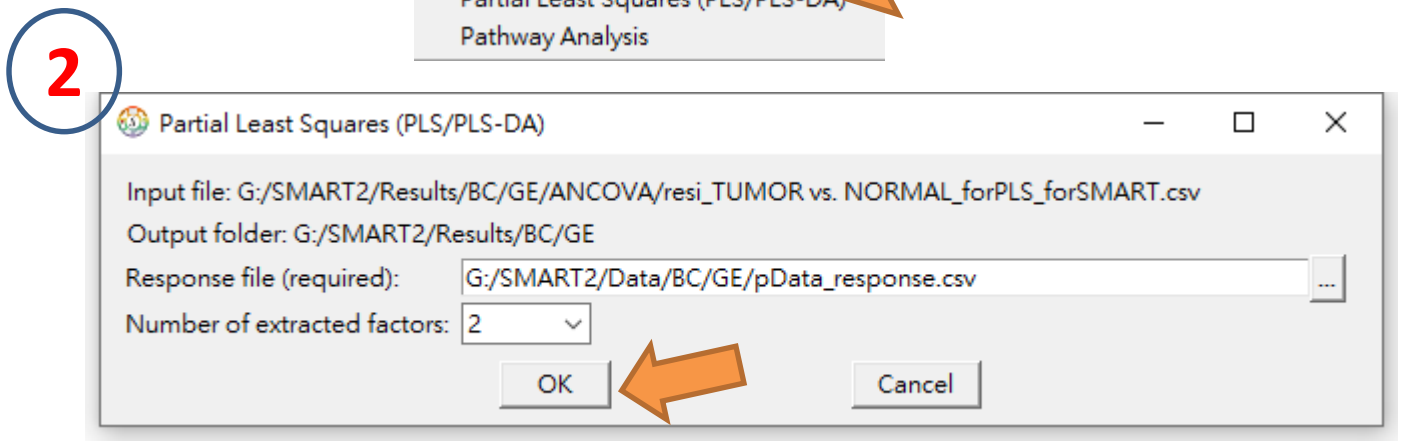
Residual file

Peak_In	mz	Ret	LHC3265	LHC3630	LHC4662	LHC5838	LHC6065
1	1	1	-0.4827	1.1505	-0.1989	-0.2741	-0.0407
2	2	2	-0.3548	0.2817	-0.2449	0.4087	0.2811
3	3	3	-0.8935	0.9525	1.1760	0.4201	0.2843
4	4	4	1.9564	0.0220	3.1320	-0.6095	2.8053
5	5	5	-0.3283	-0.1210	-0.4228	0.0158	-0.1222
6	6	6	0.1040	-0.1262	0.1050	-0.2424	0.0600
7	7	7	0.1356	-0.7696	1.5441	-0.1083	0.7173
8	8	8	0.1454	-0.1397	0.1416	-0.2649	0.0692
9	9	9	0.3133	-0.1251	0.1743	-0.2801	-0.0094

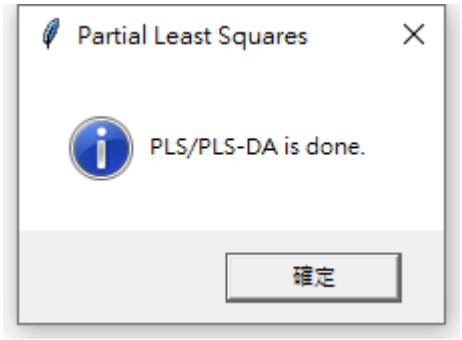
Statistical analysis (PLS/PLS-DA)

- PLS-DA
  - Input: Residuals from ANCOVA






```
PLS/PLS-DA - Please wait for PLS/PLS-DA processing 0 PLS-DA
108 samples x 20254 variables and 1 response
standard scaling of predictors and response(s)
      R2X(cum) R2Y(cum) Q2(cum) RMSEE pre ort pR2Y  pQ2
Total    0.237    0.685   0.539 0.282   2   0 0.05 0.05
```



Residual file

Peak_In	mz	Ret	LHC3265	LHC3630	LHC4662	LHC5838	LHC6065
1	1	1	-0.4827	1.1505	-0.1989	-0.2741	-0.0407
2	2	2	-0.3548	0.2817	-0.2449	0.4087	0.2811
3	3	3	-0.8935	0.9525	1.1760	0.4201	0.2843
4	4	4	1.9564	0.0220	3.1320	-0.6095	2.8053
5	5	5	-0.3283	-0.1210	-0.4228	0.0158	-0.1222
6	6	6	0.1040	-0.1262	0.1050	-0.2424	0.0600
7	7	7	0.1356	-0.7696	1.5441	-0.1083	0.7173
8	8	8	0.1454	-0.1397	0.1416	-0.2649	0.0692
9	9	9	0.3133	-0.1251	0.1743	-0.2801	-0.0094

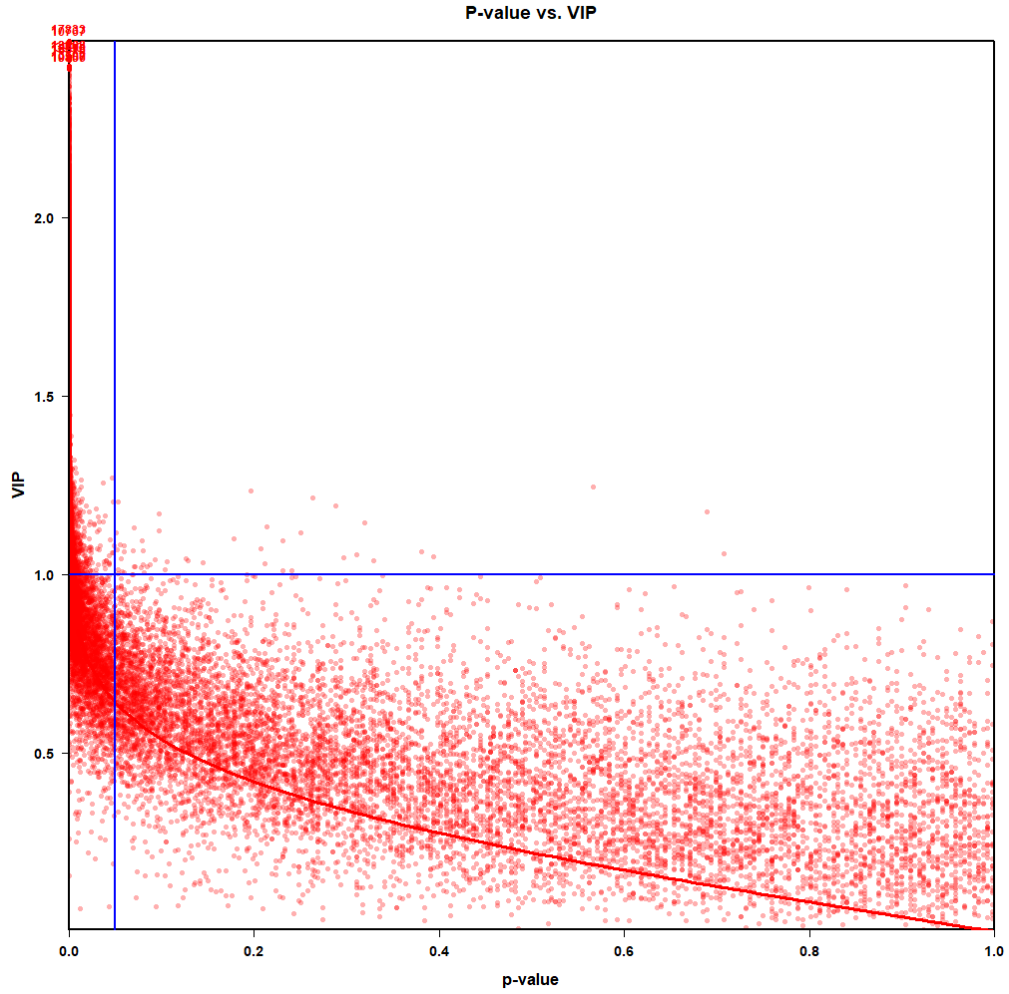
Response file

SampleID	DGroup
LHC3265	NORMAL
LHC3630	NORMAL
LHC3660	NORMAL
LHC4662	NORMAL
LHC3264	TUMOR
LHC3631	TUMOR
LHC3661	TUMOR
LHC4333	TUMOR

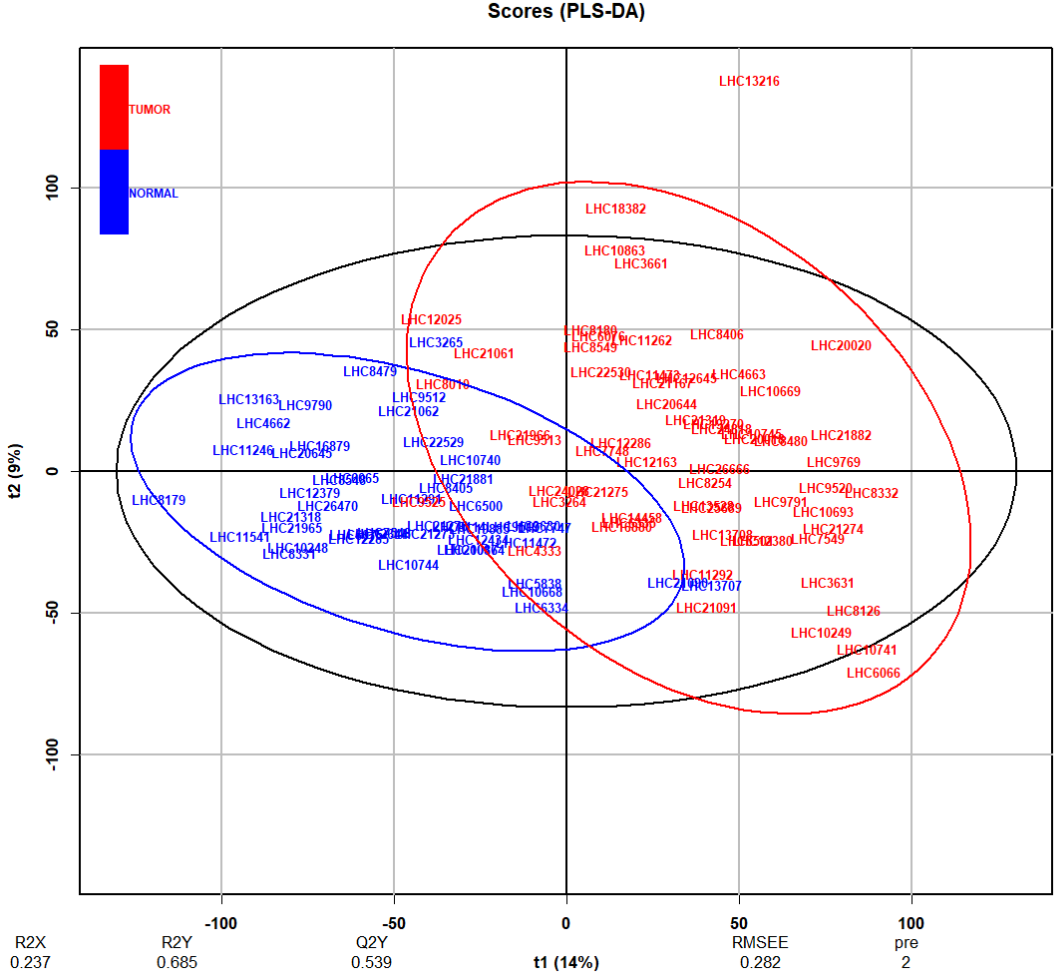
Statistical analysis (PLS/PLS-DA)

- PLS/PLS-DA output (figures)

outpath\PLS-DA\VIPvalues.png



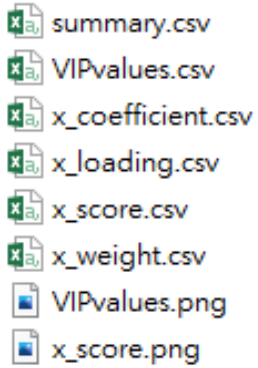
outpath\PLS-DA\x_score.png



Statistical analysis (PLS/PLS-DA)

- PLS/PLS-DA output (numerical results)

outpath\PLS-DA\



summary.csv

Variation explanation										
R2X	R2X(cum)	R2Y	R2Y(cum)	Q2	Q2(cum)	Signif.	Iter.	RMSEE	pR2Y	pQ2
0.146	0.146	0.589	0.589	0.549	0.549	R1	1			
0.1	0.246	0.0905	0.679	0.0153	0.556	R1	1	0.285	0.05	0.05

VIPvalues.csv

Peak_	mz	Ret_ti	VIP	P-value
1	1	1	0.848680576	0.051310251
2	2	2	0.957861769	0.01262452
3	3	3	0.250738031	0.9629312
4	4	4	2.224399175	3.21E-12
5	5	5	0.60991584	0.062585973

x_score.csv

Sample	p1	p2
LHC10248	-79.65393937	-9.205999313
LHC10249	69.64932109	-22.63991244
LHC10668	-19.95370836	-34.39035425
LHC10669	48.6971083	45.06335989
LHC10693	83.12152012	-30.24721724
LHC10740	-41.28603181	12.68094071
LHC10741	71.66903023	-45.54072944

*Integrative **omics** pathway analysis (IOPA)*

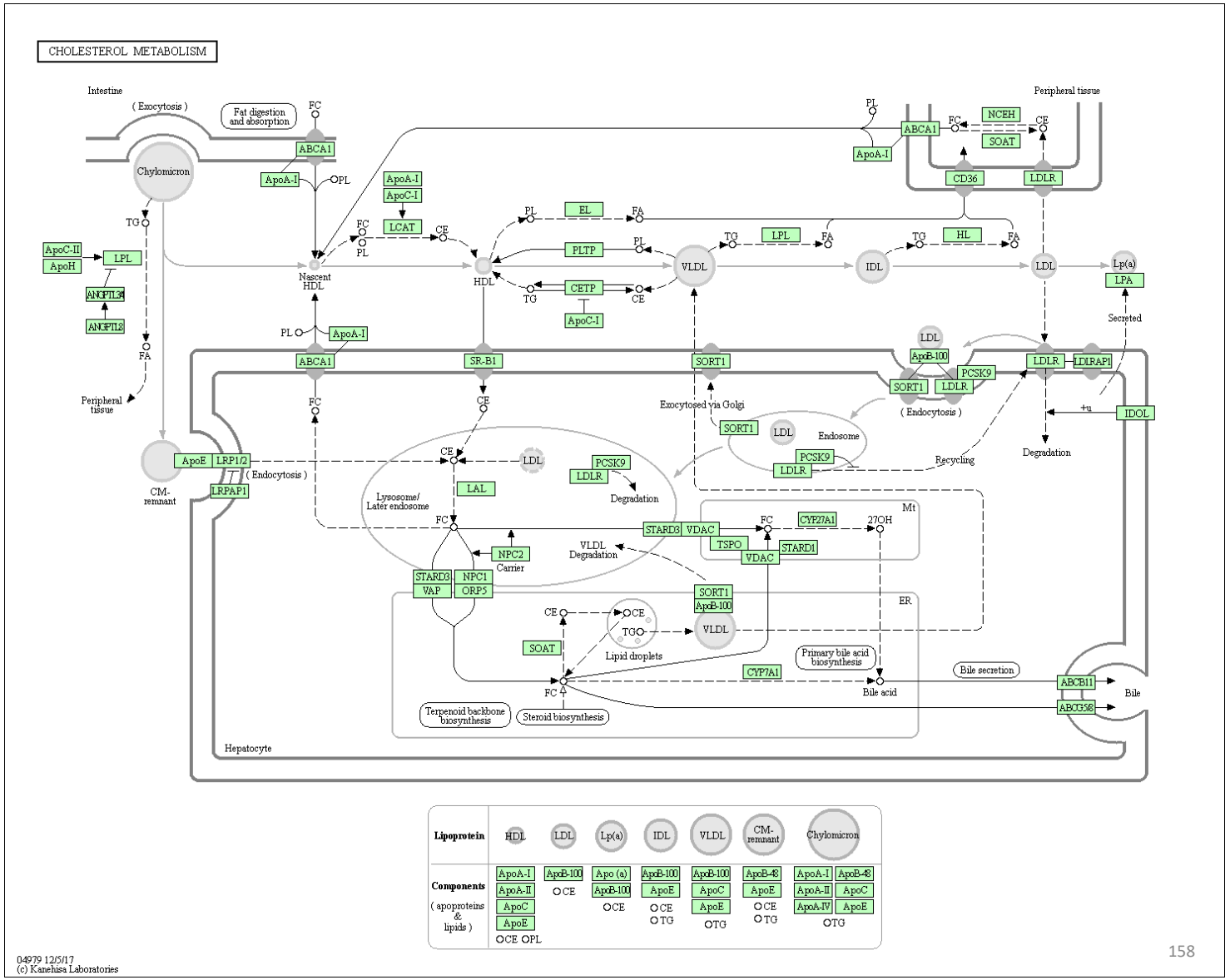
STATISTICAL ANALYSIS (3)

Integrative ***omics*** pathway analysis (IOPA)

- Multi-omics data
  - Gene expression
  - Metabolomics
- Association analysis
  - ANCOVA
    - Fold-change(beta value) and p-value
- Pathway database
  - KEGG
- Statistical methods
  - ORA (Over-representation Analysis)
  - SPIA (Signaling Pathway Impact Analysis)

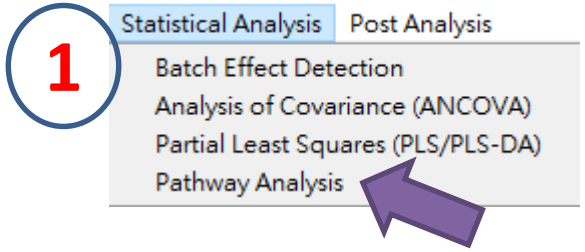
$$PF(g_i) = \Delta E(g_i) + \sum_{j=1}^{n} \beta_{ij} \cdot \frac{PF(g_j)}{N_{ds}(g_j)}$$

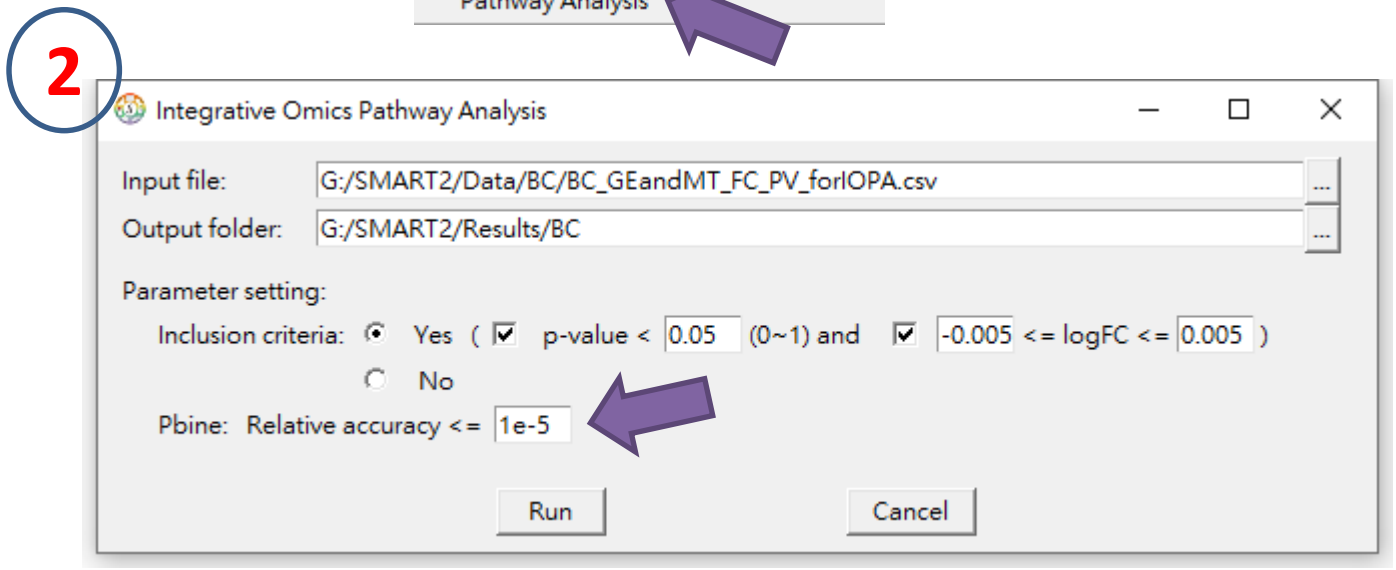
- P-value combination
  - Fisher
  - Pbine (Yin-Chun Lin et. al., 2022)



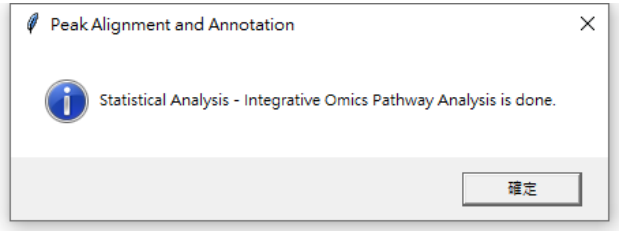
Statistical analysis (IOPA)

- IOPA
  - Input: Fold-changes and p-values from ANCOVA






```
Done pathway 301 : Diabetic cardiomyopathy..
Done pathway 302 : Viral myocarditis..
Done pathway 303 : Lipid and atherosclerosis..
Done pathway 304 : Fluid shear stress and atheros..
Post analysis and graph plotting
Progress 10.2
Progress 20.39
```



Input file

beta value

raw p-value

Gene ID

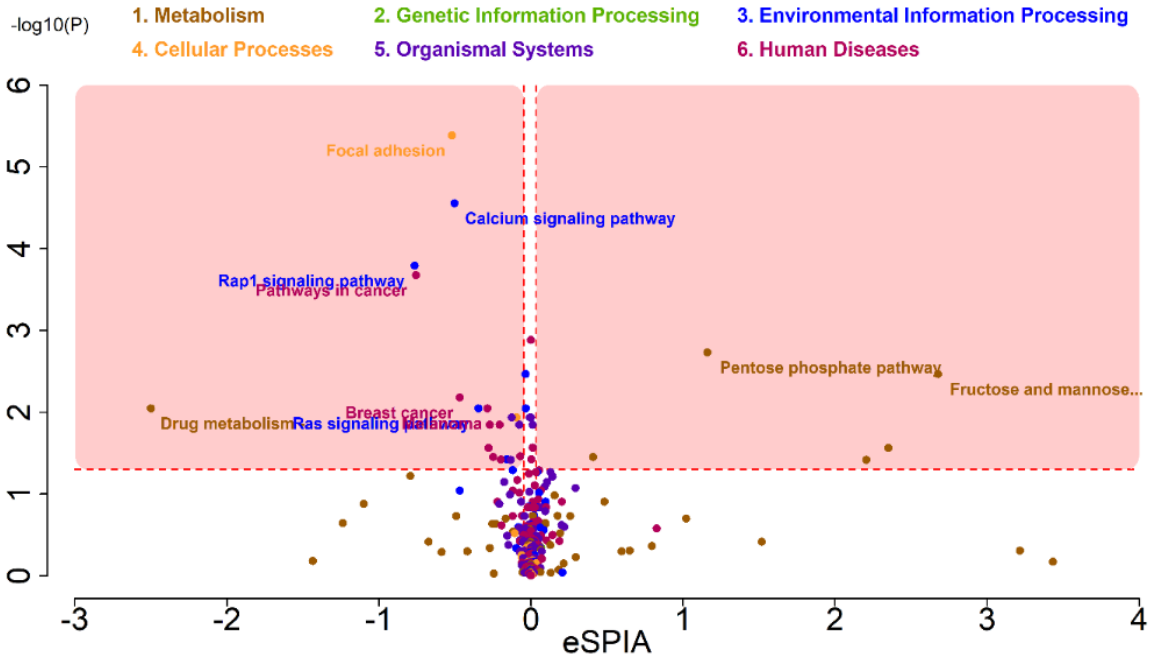
Compound ID

ID	TUMOR_NORMAL_logFC	TUMOR_NORMAL_pv
1	-0.00110327	0.071059437
2	-0.000916425	7.18E-05
9	0.015210523	0.000932944
10	0.000653423	0.630181778
12	-0.012109371	0.012604048
13	-0.003937972	2.75E-08
14	-0.002344647	0.864390741
15	-8.70E-06	0.609908795
16	-0.000819847	0.006943451
18	0.008067477	0.05162788
C00064	0.000881739	0.031417287
C00065	0.006986954	3.65E-11
C00072	0.013958216	1.24E-07
C00073	0.001264568	1.42E-07
C00074	0.013968516	1.72E-06
C00077	-0.000517312	2.16E-07
C00078	0.0037256	2.34E-10
C00079	0.004466391	4.24E-07
C00082	0.004255505	2.10E-06
C00086	0.01605101	6.20E-08

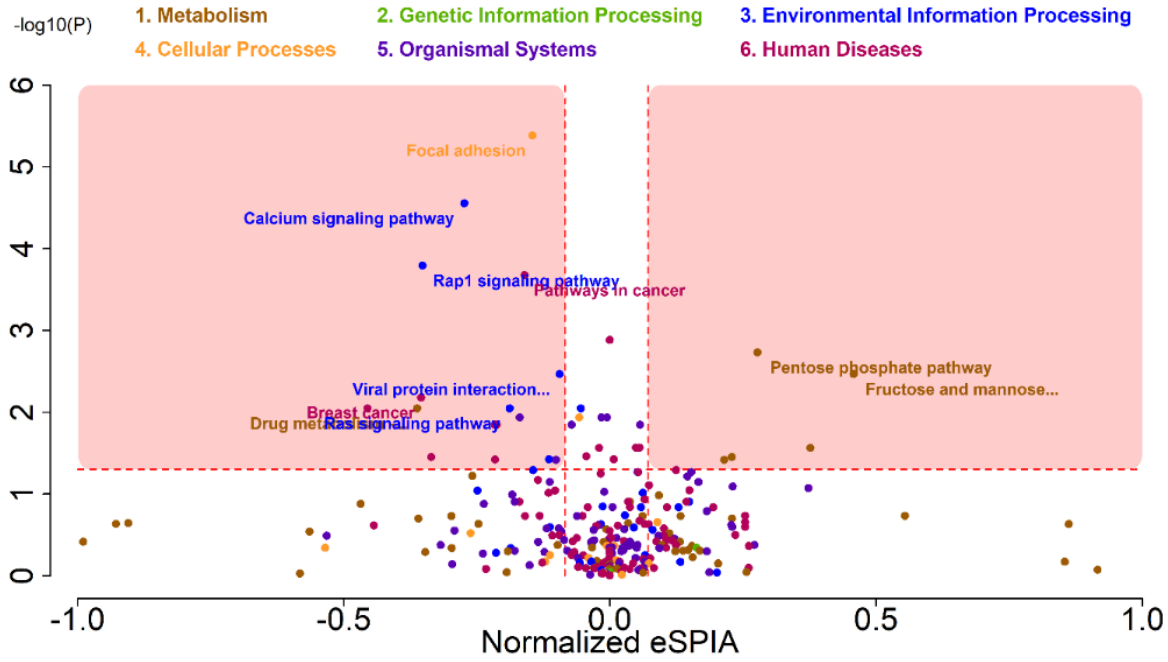
Statistical analysis (IOPA)

- IOPA output (figures)

outpath\IOPA\Pathway analysis_volcano.tiff



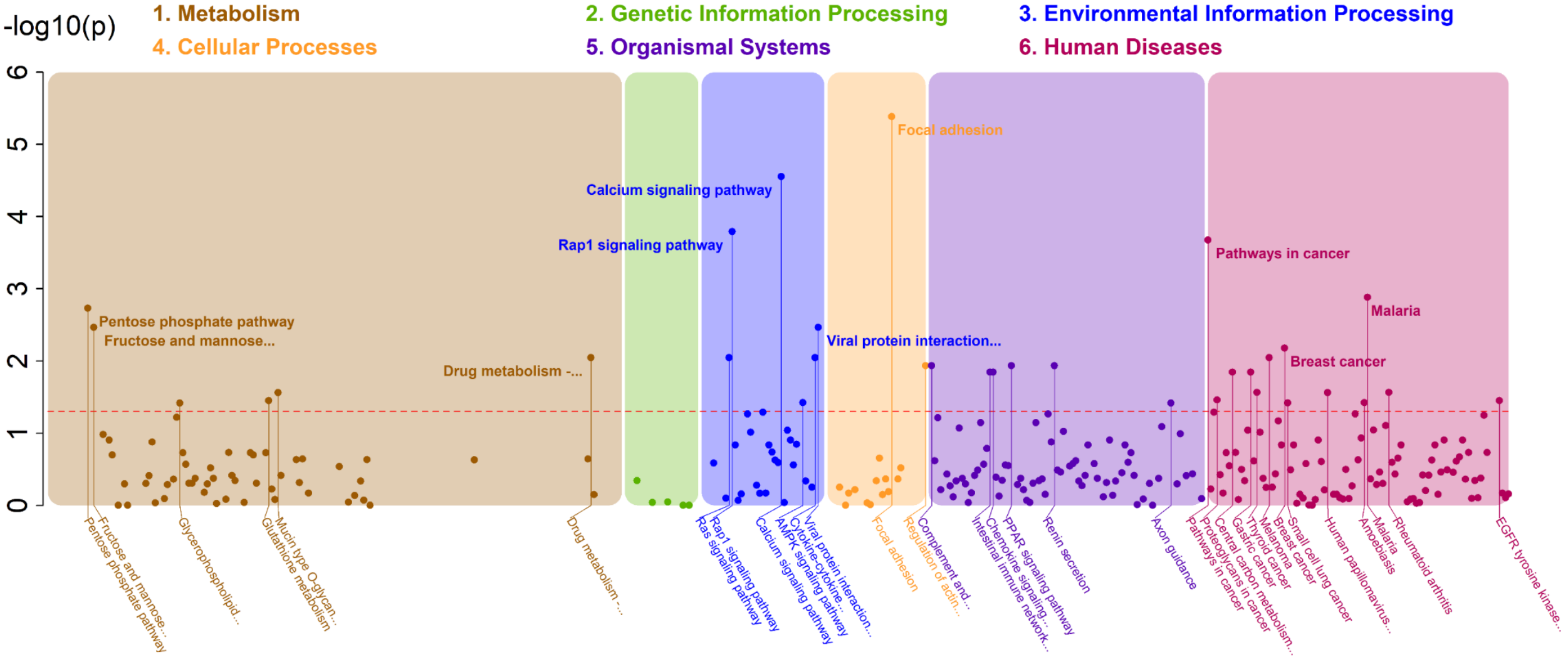
outpath\IOPA\Pathway analysis_volcano_normalized.tiff



Statistical analysis (IOPA)

- IOPA output (figures)

outpath\IOPA\Pathway analysis_manhattan.tiff



Statistical analysis (IOPA)

- IOPA output (numerical results)

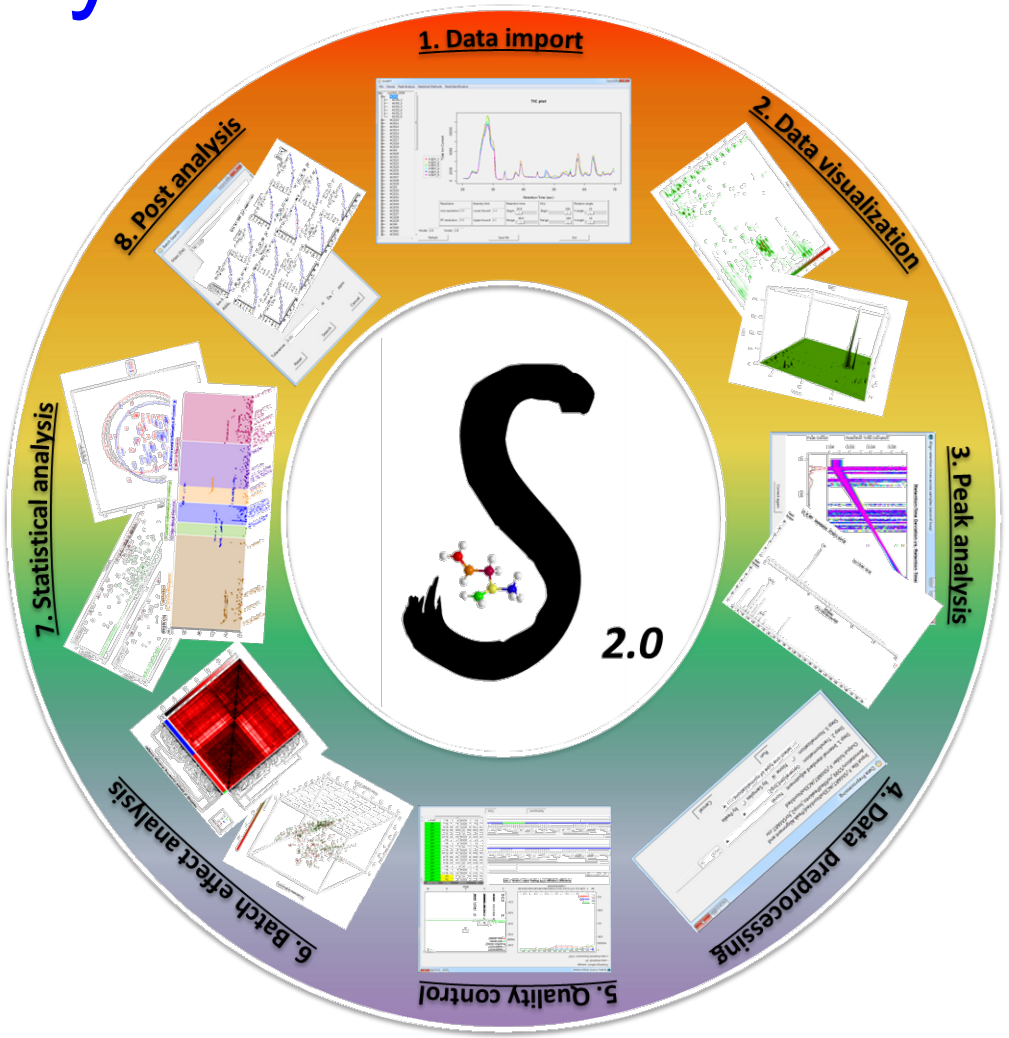
outpath\IOPA\

IOPA_fcl_fcu_pverr1e-05.csv

Name	ID	G_size	deG_n	G_pORA	C_size	deC_n	C_pORA	P_size	deP_n	pORA	pORA_Bo	pORA_FI	raw_eSPI	eSPIA	eSPIA_nc	peSPIA	peSPIA_I	peSPIA_I	pFisher	pFisher_I	pFisher_I	pPbine	pPbine_B	pPbine_F	Status	KEGGLIN
Focal adhesion	4510	199	33	2.51E-08	0	0	1	199	33	9.64E-08	2.62E-05	1.31E-05	-0.52457	-0.51835	-0.14581	0.002	0.542	0.1355	4.51E-09	1.23E-06	6.13E-07	2.22E-08	6.00E-06	6.00E-06	Inhibited	http://www
Calcium signaling pathway	4020	239	28	0.000216	0	0	1	239	28	0.000539	0.146561	0.006979	-0.50532	-0.50228	-0.27379	5.00E-06	0.001355	0.000678	5.59E-08	1.52E-05	5.06E-06	1.71E-07	4.62E-05	2.31E-05	Inhibited	http://www
Rap1 signaling pathway	4015	209	30	2.52E-06	0	0	1	209	30	7.81E-06	0.002123	0.000425	-0.77625	-0.7646	-0.35228	0.004	1	0.180667	5.71E-07	0.000155	3.88E-05	1.78E-06	0.00048	0.00016	Inhibited	http://www
Pathways in cancer	5200	529	55	1.01E-05	4	0	1	533	55	6.00E-05	0.016323	0.001814	-0.7608	-0.75518	-0.16013	0.001	0.271	0.090333	1.06E-06	0.000288	5.76E-05	3.13E-06	0.000845	0.000211	Inhibited	http://www
Malaria	5144	49	14	3.35E-07	0	0	1	49	14	6.67E-07	0.000181	6.05E-05	0	0	0	1	1	1	1.02E-05	0.002762	0.00046	2.43E-05	0.006569	0.001314	Inhibited	http://www
Pentose phosphate pathway	30	29	5	0.022018	8	6	0.020092	37	11	6.89E-06	0.001874	0.000425	1.132867	1.160247	0.277502	0.182	1	0.679356	1.83E-05	0.004977	0.000711	4.15E-05	0.011198	0.001866	Activated	http://www
Fructose and mannose metab	51	33	4	0.114111	7	6	0.006934	40	10	9.35E-05	0.025419	0.002118	2.637291	2.677882	0.457923	0.039	1	0.529095	4.93E-05	0.013405	0.001489	0.000101	0.027249	0.003406	Activated	http://www
Viral protein interaction with	4061	96	18	6.36E-06	0	0	1	96	18	1.38E-05	0.003767	0.000628	-0.03463	-0.03399	-0.09449	0.251	1	0.738832	4.72E-05	0.012831	0.001489	9.69E-05	0.02617	0.003406	Inhibited	http://www
Breast cancer	5224	147	18	0.001661	0	0	1	147	18	0.003078	0.837214	0.024223	-0.46963	-0.46674	-0.35471	0.003	0.813	0.1626	0.000116	0.031628	0.003163	0.000222	0.059848	0.00665	Inhibited	http://www
Drug metabolism - cytochror	982	59	12	9.63E-05	0	0	1	59	12	0.000165	0.044868	0.003205	-2.50511	-2.49872	-0.36212	0.12	1	0.616306	0.000234	0.063695	0.004714	0.000419	0.113101	0.008986	Inhibited	http://www
Ras signaling pathway	4014	232	25	0.001543	0	0	1	232	25	0.003295	0.896239	0.024223	-0.35004	-0.34545	-0.18837	0.006	1	0.20325	0.000234	0.063622	0.004714	0.000418	0.112984	0.008986	Inhibited	http://www
Cytokine-cytokine receptor i	4060	287	35	1.60E-05	0	0	1	287	35	5.18E-05	0.0141	0.001762	-0.03616	-0.03346	-0.05458	0.397	1	0.879675	0.000243	0.066002	0.004714	0.000433	0.116817	0.008986	Inhibited	http://www
Melanoma	5218	72	11	0.002296	0	0	1	72	11	0.003563	0.969157	0.02485	-0.29144	-0.28558	-0.45544	0.005	1	0.193571	0.000213	0.057837	0.004714	0.000384	0.103609	0.008986	Inhibited	http://www
PPAR signaling pathway	3320	72	14	4.38E-05	0	0	1	72	14	8.10E-05	0.022041	0.002118	-0.00688	-0.00688	-0.01593	0.44	1	0.889851	0.000401	0.109023	0.00647	0.000683	0.184344	0.011614	Inhibited	http://www
Regulation of actin cytoskele	4810	215	28	3.39E-05	0	0	1	215	28	9.12E-05	0.024815	0.002118	-0.10275	-0.09474	-0.05758	0.377	1	0.879675	0.000388	0.105507	0.00647	0.000663	0.178929	0.011614	Inhibited	http://www
Renin secretion	4924	69	12	0.00045	2	1	0.566055	71	13	0.000271	0.073626	0.004331	-0.12699	-0.12576	-0.1694	0.133	1	0.616306	0.000404	0.109986	0.00647	0.000688	0.185824	0.011614	Inhibited	http://www
Complement and coagulation	4610	85	16	1.95E-05	0	0	1	85	16	3.90E-05	0.010619	0.001517	-0.00525	0.000728	-0.00529	0.996	1	1	0.000434	0.117978	0.006554	0.000734	0.198064	0.011651	Activated	http://www
Chemokine signaling pathwa	4062	188	25	6.20E-05	0	0	1	188	25	0.000152	0.041463	0.003189	-0.08032	-0.0769	-0.07157	0.389	1	0.879675	0.000636	0.173113	0.008459	0.00104	0.280745	0.014279	Inhibited	http://www
Intestinal immune network fo	4672	45	10	0.000168	0	0	1	45	10	0.000268	0.072787	0.004331	0.011799	0.011799	0.056716	0.24	1	0.738832	0.000684	0.186098	0.008459	0.001111	0.299849	0.014279	Activated	http://www

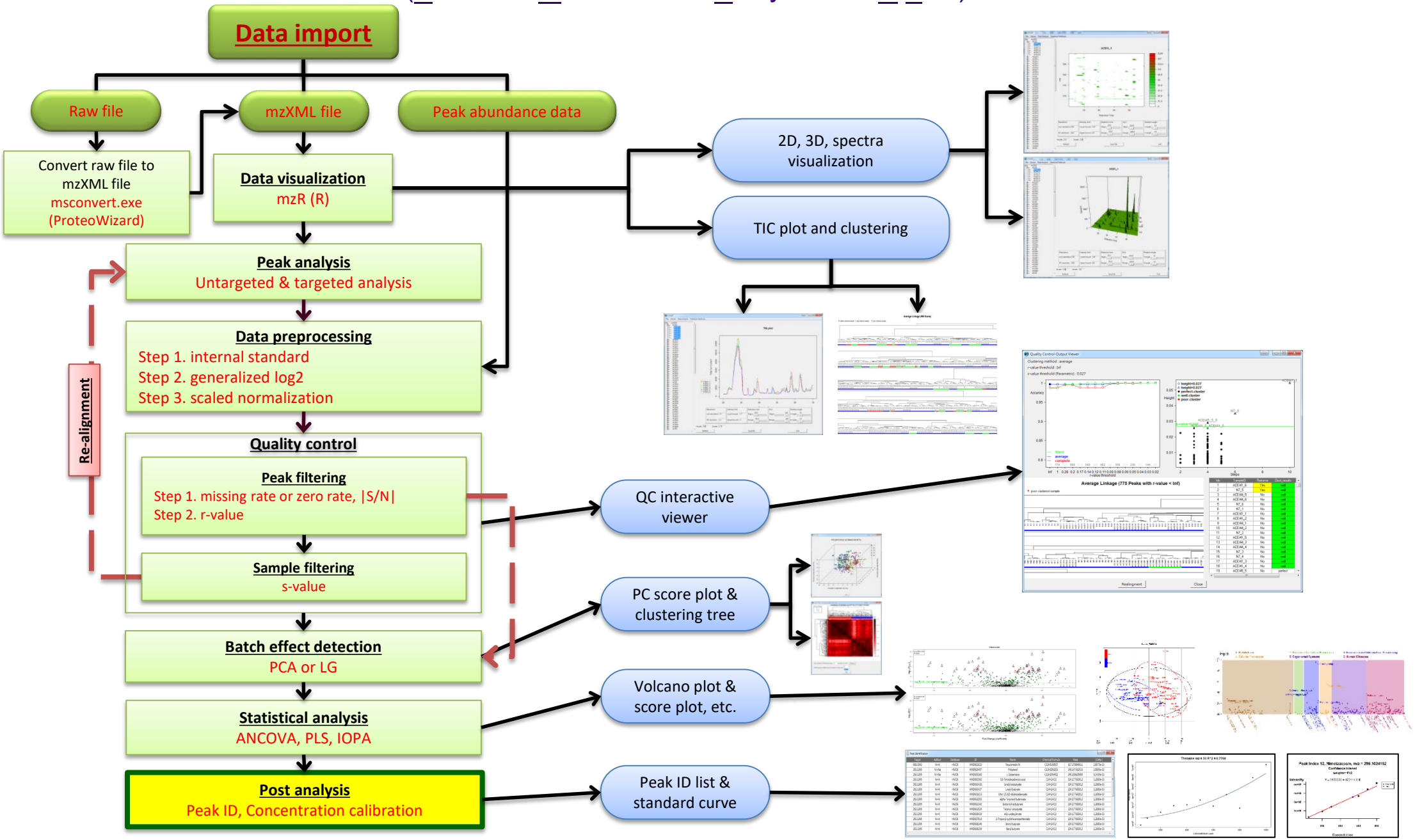
Statistical Metabolomics Analysis – an R Tool

8. Post analysis



Workflow

SMART
(Statistical Metabolomics Analysis - an R Tool)

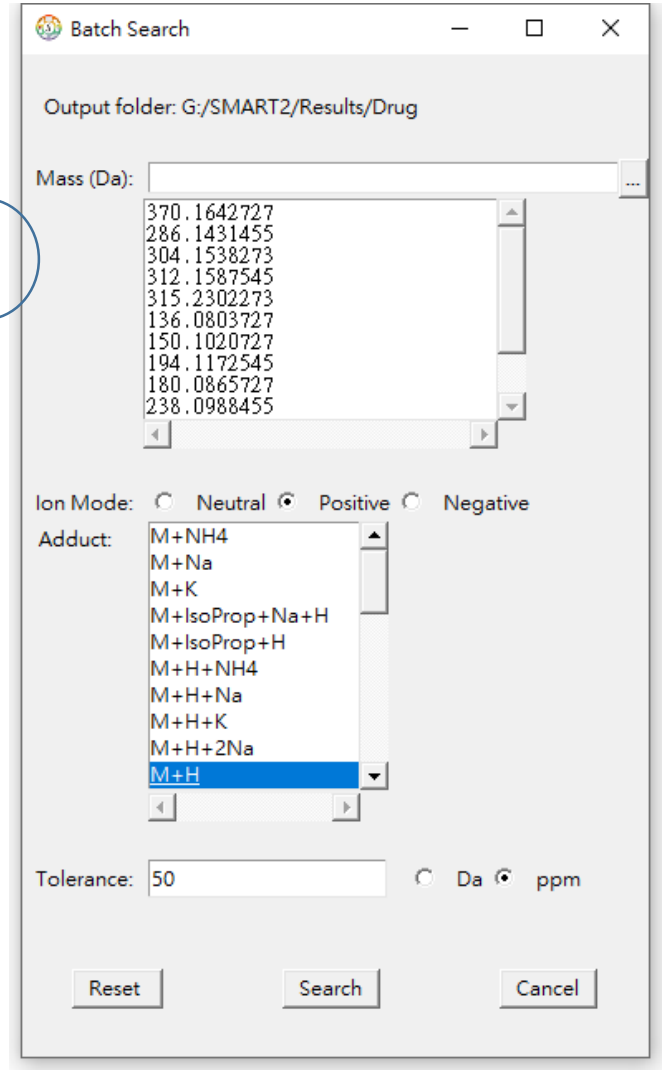


Peak identification

POST ANALYSIS (1)

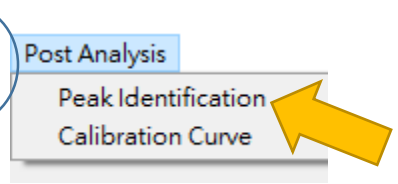
Post analysis (Peak identification)

- Input m/z



2

1



ANCOVA

Peak_Index	mz	Ret_time.sec	ACEi_NonMed
724	512.3515	116.3137	1.67E-09
425	321.3204	189.6585	5.66E-06
162	173.1252	60.0519	6.44E-06
179	183.111	52.2998	6.44E-06
308	261.1474	53.2244	6.44E-06
532	377.1709	155.9171	6.44E-06
864	619.3877	70.1135	6.44E-06
375	304.2394	155.9289	7.49E-06
427	322.3229	189.6408	7.49E-06

3

Peak Identification

Target	Adduct	Database	ID	Name	Chemical Formula	Mass	| Delta |
656.3392	M+K	HMDB	HMDB13022	Neuromedin N	C32H51N5O7	617.378849011	2.8070e-03
263.1395	M+Na	HMDB	HMDB15407	Pirbuterol	C12H20N2O3	240.147392516	2.8895e-03
263.1395	M+Na	HMDB	HMDB39160	L-Gizzerosine	C11H20N4O2	240.158625904	8.3439e-03
263.1395	M+K	HMDB	HMDB00560	5,8-Tetradecadienoic acid	C14H24O2	224.177630012	1.2880e-03
263.1395	M+K	HMDB	HMDB30426	Linalyl isobutyrate	C14H24O2	224.177630012	1.2880e-03
263.1395	M+K	HMDB	HMDB30427	Linalyl butyrate	C14H24O2	224.177630012	1.2880e-03
263.1395	M+K	HMDB	HMDB31015	Ethyl (2E,6Z)-dodecadienoate	C14H24O2	224.177630012	1.2880e-03
263.1395	M+K	HMDB	HMDB32053	alpha-Terpineol butanoate	C14H24O2	224.177630012	1.2880e-03
263.1395	M+K	HMDB	HMDB32342	Isobornyl isobutyrate	C14H24O2	224.177630012	1.2880e-03
263.1395	M+K	HMDB	HMDB32524	Terpinyl isobutyrate	C14H24O2	224.177630012	1.2880e-03
263.1395	M+K	HMDB	HMDB36424	Allyl undecylenate	C14H24O2	224.177630012	1.2880e-03
263.1395	M+K	HMDB	HMDB37814	2-Propenyl cyclohexanepentanoate	C14H24O2	224.177630012	1.2880e-03
263.1395	M+K	HMDB	HMDB38246	Bornyl butyrate	C14H24O2	224.177630012	1.2880e-03
263.1395	M+K	HMDB	HMDB38259	Neryl butyrate	C14H24O2	224.177630012	1.2880e-03

HMDB & MassBank

Post analysis (Peak identification)

- Peak ID output (numerical results)

outpath\Peak Identification\

peakID_search_results_2023-03-31 18-09-07.csv

Target	Adduct.Model	DB	ID	Name	Chem.Formu	Mass	abs.delta
512.3515	M	HMDB	HMDB31879	11a,12a-Epoxy-3b-hydroxy-28,13-oleananolide 3-acetate	C32H48O5	512.3501746	0.001325354
512.3515	M	HMDB	HMDB34646	3beta-Acetoxy-12-oxo-28,13beta-oleananolide	C32H48O5	512.3501746	0.001325354
512.3515	M	HMDB	HMDB35313	Ganoderic acid S	C32H48O5	512.3501746	0.001325354
512.3515	M	HMDB	HMDB35330	Ganoderic acid Mf	C32H48O5	512.3501746	0.001325354
512.3515	M	HMDB	HMDB35332	Ganoderic acid X	C32H48O5	512.3501746	0.001325354
512.3515	M	HMDB	HMDB36672	3alpha-Acetoxy-11-keto-beta-boswellic acid	C32H48O5	512.3501746	0.001325354
173.1252	M	HMDB	HMDB59607	apo-[3-methylcrotonoyl-CoA:carbon-dioxide ligase (ADP-forming)]	C7H15N3O2	173.1164267	0.008773261
183.111	M	HMDB	HMDB40067	4-Butyl-5-propylthiazole	C10H17NS	183.1081702	0.002829761
183.111	M	HMDB	HMDB40083	2,4-Dimethyl-5-pentylthiazole	C10H17NS	183.1081702	0.002829761
183.111	M	HMDB	HMDB40084	4,5-Dimethyl-2-pentylthiazole	C10H17NS	183.1081702	0.002829761

Post analysis (Peak identification)

- Peak ID output (numerical results)

outpath\Peak Identification\

peakID_search_results_2023-10-18 23-18-16_100ppm.csv

Target	Adduct	DB	DB_ID	Name	Chem.Formu	Mass	KEGG_ID	abs.delta
370.1643	M+H	HMDB	HMDB0014038	N-Deisopropyl-fluvastatin	C21H20FNO4	369.1376		0.01936
370.1643	M+H	HMDB	HMDB0015288	Trimetrexate	C19H23N5O3	369.1801	C11154	0.023093
370.1643	M+H	HMDB	HMDB0015633	Amisulpride	C17H27N3O4S	369.1722		0.01523
370.1643	M+H	HMDB	HMDB0029360	Isorheagenine	C20H19NO6	369.1212		0.035759
370.1643	M+H	HMDB	HMDB0029401	Ochratoxin B	C20H19NO6	369.1212		0.035759
370.1643	M+H	HMDB	HMDB0031947	Niazicinin	C17H23NO8	369.1424		0.01463
286.1431	M+H	HMDB	HMDB0002171	Glycylprolylhydroxyproline	C12H19N3O5	285.1325		0.003399
286.1431	M+H	HMDB	HMDB0014440	Morphine	C17H19NO3	285.1365	C01516	0.000624
286.1431	M+H	HMDB	HMDB0014472	Hydromorphone	C17H19NO3	285.1365	C07042	0.000624
286.1431	M+H	HMDB	HMDB0015692	Isothipendyl	C16H19N3S	285.13		0.005901
286.1431	M+H	HMDB	HMDB0029377	Piperine	C17H19NO3	285.1365	C03882	0.000624
286.1431	M+H	HMDB	HMDB0030256	Erysopine	C17H19NO3	285.1365		0.000624
286.1431	M+H	HMDB	HMDB0032961	Secodemethylclausenamide	C17H19NO3	285.1365		0.000624

Concentration calibration

- *Calibration curve construction*
- *Concentration calculation*

POST ANALYSIS (2)

Calibration curve construction

Calibration curve construction

- Input file : peak abundance data
  - Compound name and peak abundance in different concentration levels (required)
  - Sample name: D1_50_MS1 …
    - Rule: sampledID_conc._MSx
- Method :
  - Regression (linear or quadratic)
  - Weight (1, 1/x, $1/x^2$)
  - Outlier detection
    - None, Cook's D, 95% CI, Deviation
- Output file :
  - Per drug : PID1_Heroin_mz370.1642727_lm.tiff
  - Summarized files
    - Calibration-curve_info.csv
    - Peak_abundance_S10_P12.csv

Input file: peak abundance data (from Peak analysis)
-- standard compounds

Peak_Index	Name	mz_est	Ret_Time_est.sec	D1_50_MS1	D10_900_MS1	D11_1000_MS1	D2_100_MS1	D3_200_MS1	D4_300_MS1	D
1	Heroin	370.1643	685.3724	20454233	4.85E+08	5.37E+08	33246830	65461545	1.08E+08	
2	Morphine	286.1431	146.2696	41215408	5.53E+08	6.23E+08	55632394	1.11E+08	1.59E+08	
3	Cocaine	304.1538	701.136	67391965	1.91E+09	2.16E+09	1.09E+08	2.8E+08	4.54E+08	
4	Thebaine	312.1588	678.8176	13918057	6.54E+08	7.9E+08	22156818	57057964	1.12E+08	
5	delta9-THC	315.2302	1513.319	584689	12324894	10750614	1040859	3017275	4419165	
6	Amphetamine	136.0804	419.1944	28233349	38237498	37707935	30552141	36745807	37802678	
7	MA	150.1021	496.8365	88366857	4.84E+08	5.19E+08	1.14E+08	1.84E+08	2.48E+08	
8	MDMA	194.1173	538.9991	57005641	1.36E+09	1.61E+09	77917415	1.73E+08	3.32E+08	
9	Love Drug	180.0866	505.6407	2977060	32927321	40623589	3508615	5215888	7552811	
10	Ketamine	238.0988	599.9313	65361457	1.9E+09	2.12E+09	1.1E+08	3.04E+08	5.13E+08	
11	FM2	314.093	1044.274	23235201	4.47E+08	5.37E+08	37148568	83573478	1.28E+08	
12	Nimetazepam	296.1024	1040.522	27111684	5.03E+08	5.88E+08	41037216	91247724	1.44E+08	

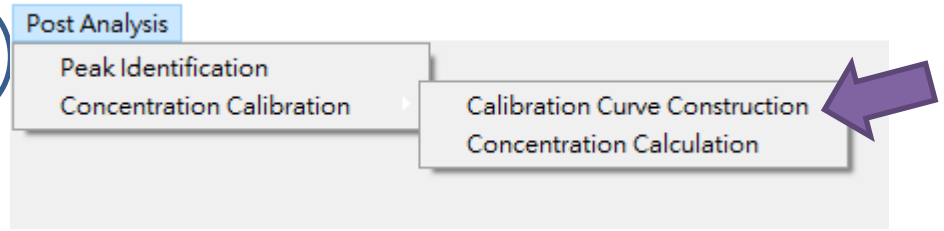
Output file: Calibration-curve_info.csv

Peak_Index	Name	mz	Ret_time.sec	model	weights	Original	concentration	theta1
1	Heroin	370.1643	685.3724	linear	unweighted		50;100;200;300;400;600;700;800;900;1000	-4.9E-
1	Heroin	370.1643	685.3724	linear	1/x		50;100;200;300;400;600;700;800;900;1000	-1.5E-
1	Heroin	370.1643	685.3724	linear	1/x2		50;100;200;300;400;600;700;800;900;1000	-3676(
1	Heroin	370.1643	685.3724	quadratic	unweighted		50;100;200;300;400;600;700;800;900;1000	19931
1	Heroin	370.1643	685.3724	quadratic	1/x		50;100;200;300;400;600;700;800;900;1000	14796
1	Heroin	370.1643	685.3724	quadratic	1/x2		50;100;200;300;400;600;700;800;900;1000	11317
2	Morphine	286.1431	146.2696	linear	unweighted		50;100;200;300;400;600;700;800;900;1000	-2.2E-
2	Morphine	286.1431	146.2696	linear	1/x		50;100;200;300;400;600;700;800;900;1000	2718
2	Morphine	286.1431	146.2696	linear	1/x2		50;100;200;300;400;600;700;800;900;1000	12145
2	Morphine	286.1431	146.2696	quadratic	unweighted		50;100;200;300;400;600;700;800;900;1000	27509
2	Morphine	286.1431	146.2696	quadratic	1/x		50;100;200;300;400;600;700;800;900;1000	25624
2	Morphine	286.1431	146.2696	quadratic	1/x2		50;100;200;300;400;600;700;800;900;1000	24114

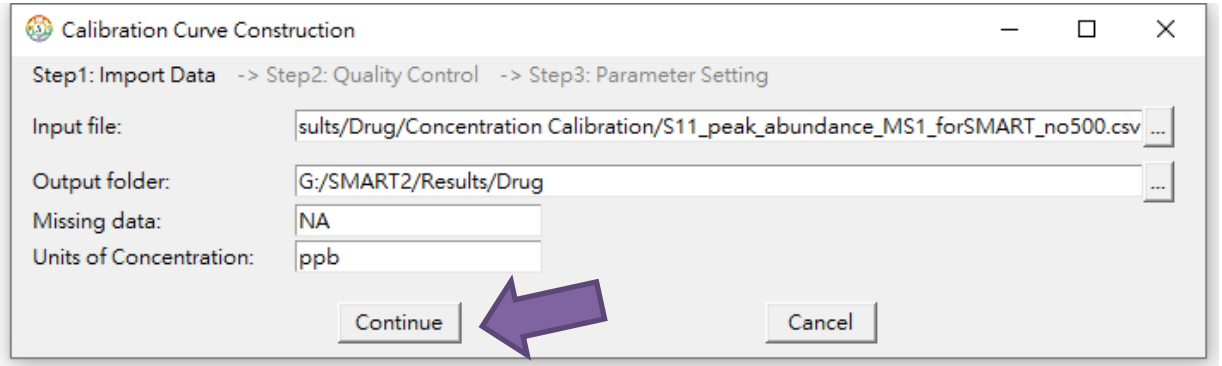
Post analysis (Concentration calibration)

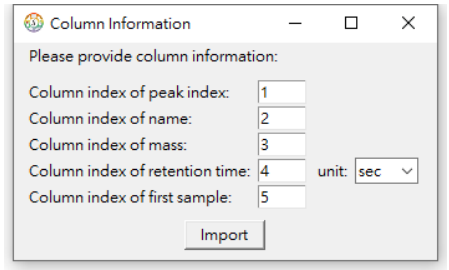
- Calibration curve construction
  - Input: Peak abundance table

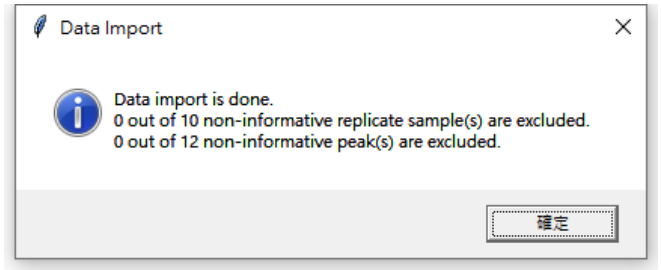
1



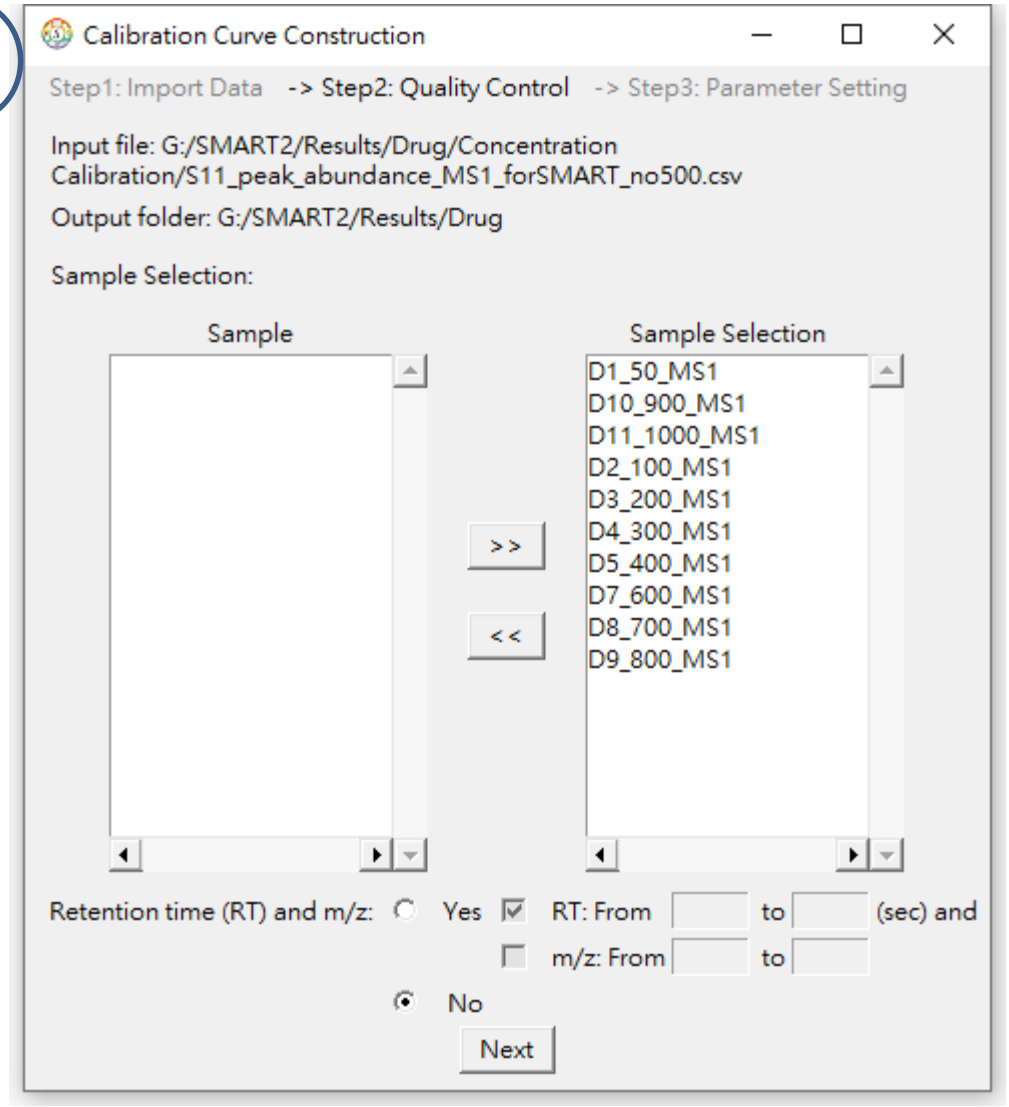
2







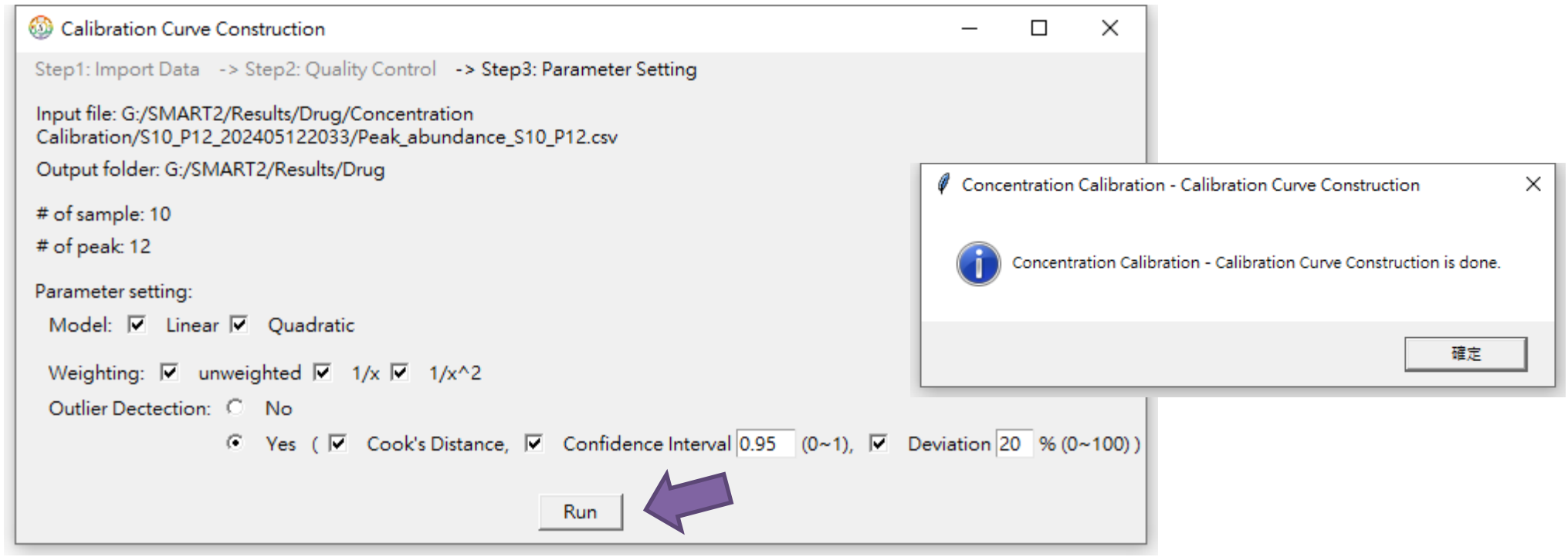
3



Post analysis (Concentration calibration)

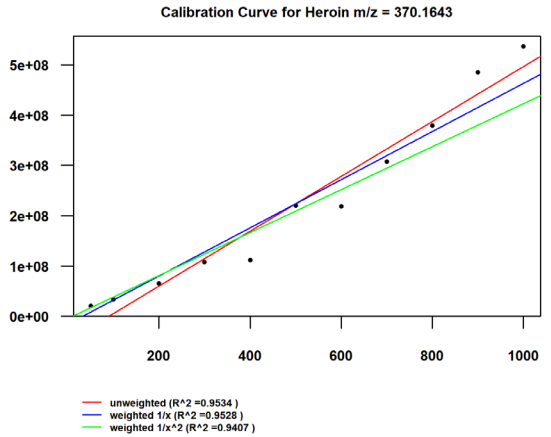
- Calibration curve construction
  - Input: Peak abundance table

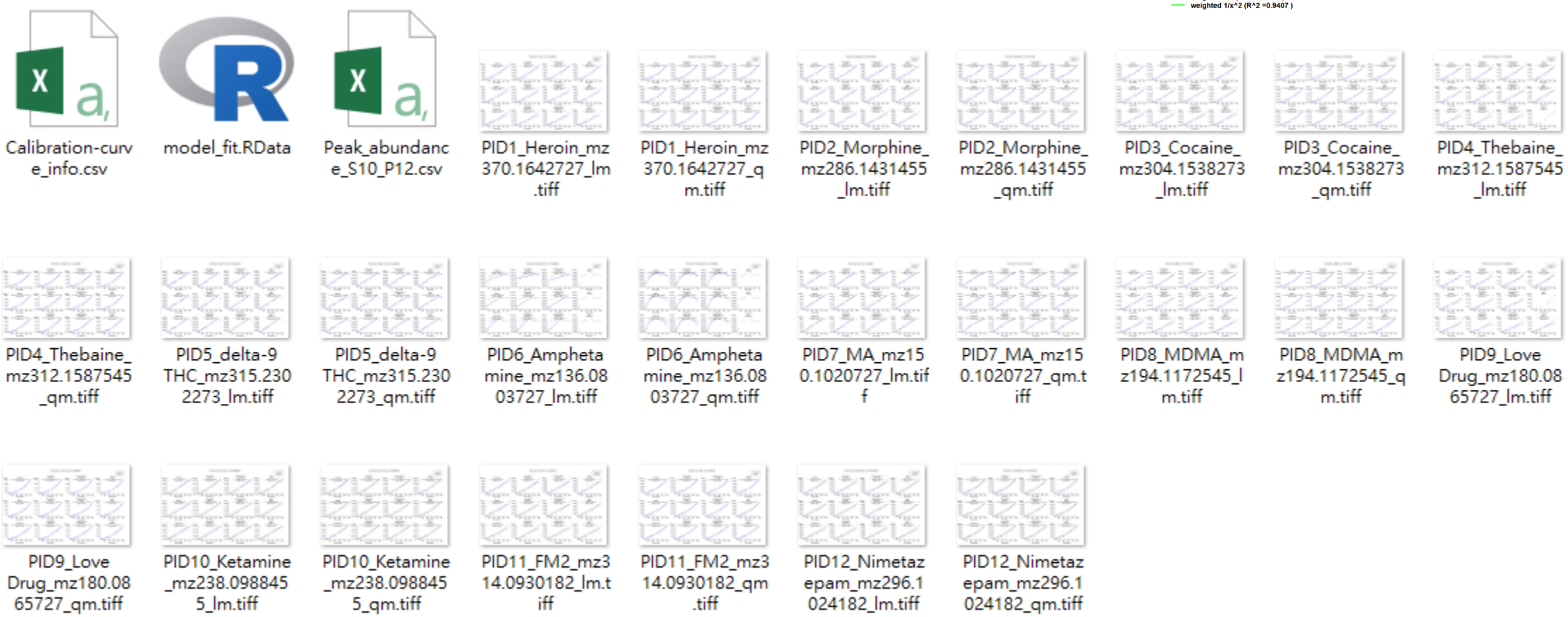
4



Calibration curve construction

- Output file :
  - Calibration curve
  - Summarized files
    - Calibration-curve_info.csv
    - Peak_abundance_S10_P12.csv





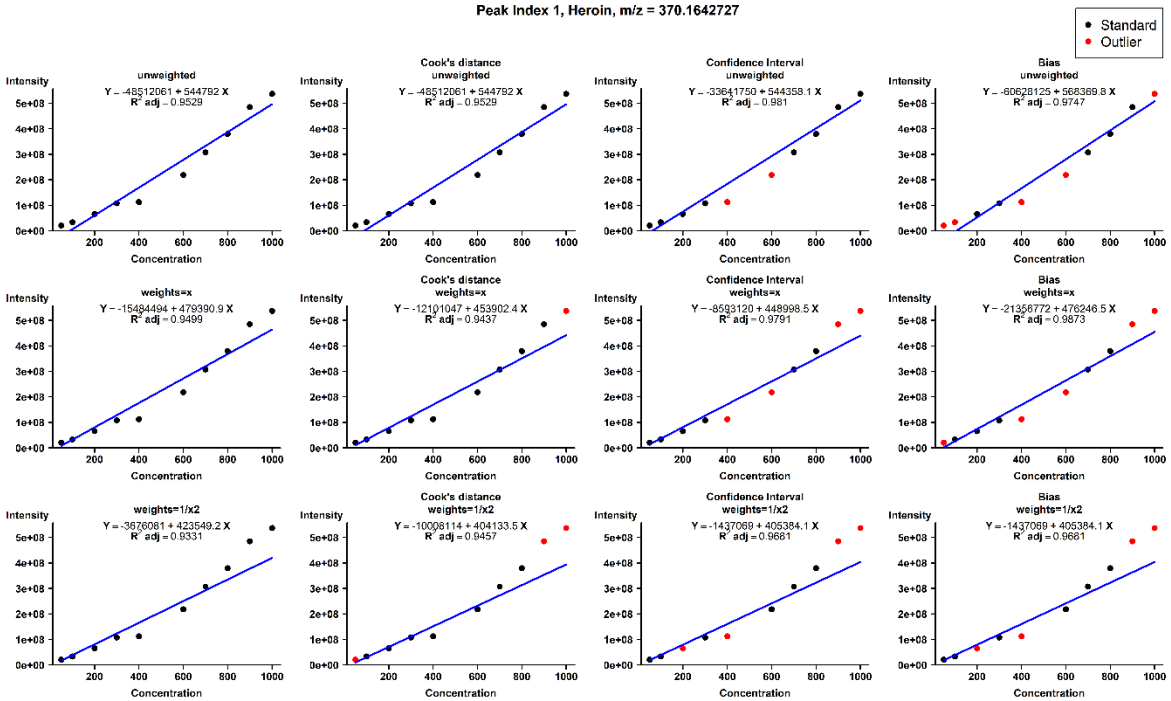
Calibration curve construction

Heroin

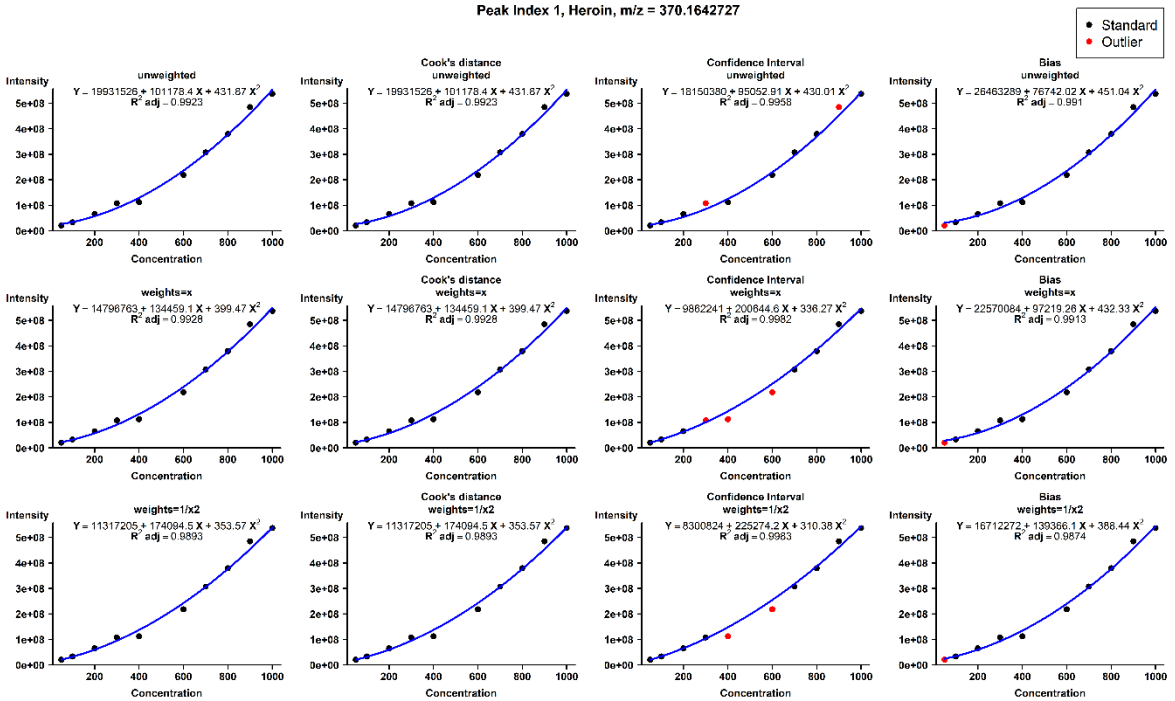
Output file :

- Calibration curve
  - PID1_Heroin_mz370.1642727_lm.tiff
  - PID1_Heroin_mz370.1642727_qm.tiff

Linear



Quadratic



Calibration curve construction

- Output file :
  - Per drug : PID1_Heroin_mz370.1642727_lm.tiff
  - Summarized files
    - Calibration-curve_info.csv
    - Peak_abundance_S10_P12.csv

Peak_Index	Name	mz	Ret_time.sec	model	weights	Original	concentration	theta1	theta2	theta3	R2	adj.R2	Cook's Distance	concentration	theta1	theta2	theta3	R2	adj.R2	Confidence Interval	concentration	theta1	theta2	theta3	R2	adj.R2	Bias	Bias_na	concentration	theta1	theta2	theta3	R2	adj.R2
1	Heroin	370.1643	685.3724	linear	unweighted		50;100;200;300;400;600;700;800;900;1000	-4.9E+07	544792		0.958136	0.952903		50;100;200;300;400;600;700;800;900;1000	-4.9E+07	544792		0.958136	0.952903		50;100;200;300;700;800;900;1000	-3.4E+07	544358.1		0.98374	0.98103		1000	200;300;700;800;900	-6.1E+07	568369.8		0.981002	0.974669
1	Heroin	370.1643	685.3724	linear	1/x		50;100;200;300;400;600;700;800;900;1000	-1.5E+07	479390.9		0.955433	0.949862		50;100;200;300;400;600;700;800;900	-1.2E+07	453902.4		0.950724	0.943685		50;100;200;300;700;800	-8593120	448998.5		0.983258	0.979073		900;1000	100;200;300;700;800	-2.1E+07	476246.5		0.990442	0.987257
1	Heroin	370.1643	685.3724	linear	1/x2		50;100;200;300;400;600;700;800;900;1000	-3676081	423549.2		0.940529	0.933096		100;200;300;400;600;700;800	-1E+07	404133.5		0.954751	0.945701		50;100;300;600;700;800	-1437069	405384.1		0.974506	0.968132		900;1000	50;100;300;600;700;800	-1437069	405384.1		0.974506	0.968132
1	Heroin	370.1643	685.3724	quadratic	unweighted		50;100;200;300;400;600;700;800;900;1000	19931526	101178.4	431.8703	0.994044	0.992342		50;100;200;300;400;600;700;800;900;1000	19931526	101178.4	431.8703	0.994044	0.992342		50;100;200;400;600;700;800;1000	18150380	95052.91	430.0097	0.997018	0.995826		50	100;200;300;400;600;700;800;900;1000	26463289	76742.02	451.0403	0.993256	0.991008
1	Heroin	370.1643	685.3724	quadratic	1/x		50;100;200;300;400;600;700;800;900;1000	14796763	134459.1	399.4706	0.994369	0.99276		50;100;200;300;400;600;700;800;900;1000	14796763	134459.1	399.4706	0.994369	0.99276		50;100;200;700;800;900;1000	9862241	200644.6	336.2698	0.9988	0.9982		50	100;200;300;400;600;700;800;900;1000	22570084	97219.26	432.3327	0.993467	0.991289
1	Heroin	370.1643	685.3724	quadratic	1/x2		50;100;200;300;400;600;700;800;900;1000	11317205	174094.5	353.5682	0.991672	0.989293		50;100;200;300;400;600;700;800;900;1000	11317205	174094.5	353.5682	0.991672	0.989293		50;100;200;300;700;800;900;1000	8300824	225274.2	310.381	0.998758	0.998261		50	100;200;300;400;600;700;800;900;1000	16712272	139366.1	388.4399	0.990517	0.987355
2	Morphine	286.1431	146.2696	linear	unweighted		50;100;200;300;400;600;700;800;900;1000	-2.2E+07	614982.2		0.980222	0.97775		50;100;200;300;400;600;700;800;900;1000	-2.2E+07	614982.2		0.980222	0.97775		50;100;200;300;700;800;900;1000	-8656085	612829.3		0.995813	0.995116		1000	200;300;600;700;800;900	-2.9E+07	620594.2		0.984702	0.980877
2	Morphine	286.1431	146.2696	linear	1/x		50;100;200;300;400;600;700;800;900;1000	2718972	565511.4		0.978751	0.976095		100;200;300;400;600;700;800;900	-9851392	574439.3		0.976138	0.972161		50;100;200;300;700;800;900;1000	3511241	588795		0.99382	0.99279		1000	100;200;300;600;700;800;900	-6890489	583750.9		0.991752	0.990102
2	Morphine	286.1431	146.2696	linear	1/x2		50;100;200;300;400;600;700;800;900;1000	12145662	520932.9		0.96943	0.965609		100;200;300;400;600;700;800	2524560	522646.5		0.974253	0.969104		50;200;300;600;700;800	14920579	514716.7		0.992002	0.990002		900;1000	50;100;200;300;600;700;800	13432841	508416.2		0.982305	0.978766
2	Morphine	286.1431	146.2696	quadratic	unweighted		50;100;200;300;400;600;700;800;900;1000	27509410	292379.9	314.0624	0.995468	0.994173		50;100;200;300;400;600;700;800;900;1000	27509410	292379.9	314.0624	0.995468	0.994173		50;100;200;300;600;700;800;900;1000	24312931	346814.2	257.3256	0.998147	0.99753		50	100;200;300;400;600;700;800;900;1000	29484511	284990.8	319.8591	0.994575	0.992767
2	Morphine	286.1431	146.2696	quadratic	1/x		50;100;200;300;400;600;700;800;900;1000	25624342	304597.9	302.1678	0.995151	0.993766		50;100;200;300;400;600;700;800;900;1000	25624342	304597.9	302.1678	0.995151	0.993766		50;100;200;300;600;700;800;900;1000	21499894	366836.4	237.6349	0.99833	0.997773		50	100;200;300;400;600;700;800;900;1000	27109026	297485.2	308.4444	0.993657	0.991542
2	Morphine	286.1431	146.2696	quadratic	1/x2		50;100;200;300;400;600;700;800;900;1000	24114845	321792.5	282.2545	0.991638	0.989249		50;100;200;300;400;600;700;800;900;1000	24114845	321792.5	282.2545	0.991638	0.989249		50;100;200;300;600;700;800;900;1000	21580848	365864.8	238.7356	0.996388	0.995184			50;100;200;300;400;600;700;800;900;1000	24114845	321792.5	282.2545	0.991638	0.989249

Calibration curve construction

- Output file :
  - Per drug : PID1_Heroin_mz370.1642727_lm.tiff
  - Summarized files
    - Calibration-curve_info.csv
    - Peak_abundance_S10_P12.csv

Peak_Index	Name	mz	Ret_time.sec	D1_50_MS1	D10_900_MS1	D11_1000_MS1	D2_100_MS1	D3_200_MS1	D4_300_MS1	D5_400_MS1	D7_600_MS1	D8_700_MS1	D9_800_MS1
1	Heroin	370.1643	685.3724	20454233	4.85E+08	5.37E+08	33246830	65461545	1.08E+08	1.12E+08	2.18E+08	3.07E+08	3.79E+08
2	Morphine	286.1431	146.2696	41215408	5.53E+08	6.23E+08	55632394	1.11E+08	1.59E+08	1.66E+08	3.05E+08	4.01E+08	4.69E+08
3	Cocaine	304.1538	701.136	67391965	1.91E+09	2.16E+09	1.09E+08	2.8E+08	4.54E+08	4.79E+08	9.95E+08	1.32E+09	1.57E+09
4	Thebaine	312.1588	678.8176	13918057	6.54E+08	7.9E+08	22156818	57057964	1.12E+08	1.14E+08	3.04E+08	4.41E+08	5.71E+08
5	delta-9 THC	315.2302	1513.319	584689	12324894	10750614	1040859	3017275	4419165	4255491	8052278	9572914	11385840
6	Amphetamine	136.0804	419.1944	28233349	38237498	37707935	30552141	36745807	37802678	38425018	39861533	37695091	38418264
7	MA	150.1021	496.8365	88366857	4.84E+08	5.19E+08	1.14E+08	1.84E+08	2.48E+08	2.57E+08	3.78E+08	4.43E+08	4.61E+08
8	MDMA	194.1173	538.9991	57005641	1.36E+09	1.61E+09	77917415	1.73E+08	3.32E+08	3.11E+08	7.22E+08	9.57E+08	1.14E+09
9	Love Drug	180.0866	505.6407	2977060	32927321	40623589	3508615	5215888	7552811	8060482	15982588	21387478	24410233
10	Ketamine	238.0988	599.9313	65361457	1.9E+09	2.12E+09	1.1E+08	3.04E+08	5.13E+08	5.38E+08	1.13E+09	1.41E+09	1.64E+09
11	FM2	314.093	1044.274	23235201	4.47E+08	5.37E+08	37148568	83573478	1.28E+08	1.31E+08	2.43E+08	3.1E+08	3.84E+08
12	Nimetazepam	296.1024	1040.522	27111684	5.03E+08	5.88E+08	41037216	91247724	1.44E+08	1.48E+08	2.67E+08	3.45E+08	4.14E+08

Concentration calculation

Concentration calculation

- Input: peak abundance data of test samples
  - Compound name and peak abundance
  - Test sample name: T1_500_MS1
    - Rule: sampledID_MSx

- Input: Model Info. of *Calibration curve construction*
  - Regression (linear or quadratic)
  - Weight (1, 1/x, $1/x^2$)
  - Outlier detection
    - Cook's D, 95% CI, Deviation
  - Rdata from *Calibration curve construction*

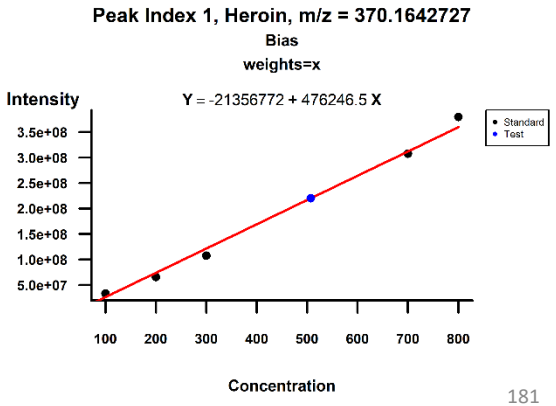
- Output file :
  - per drug : Conc.Calculation_PID1_S1_P12.tiff
  - Summarized files
    - Conc.Calculation_S1_P12.csv
    - Conc.Calculation_CI_S1_P12.csv
    - best_model.csv

Input file: peak abundance data of test samples
Model info: model_fit.RData

Peak_Index	Name	mz_est	Ret_Time_est.sec	T1_500_MS1
1	Heroin	370.1643	685.3724	2.2E+08
2	Morphine	286.1431	146.2696	2.97E+08
3	Cocaine	304.1538	701.136	9.74E+08
4	Thebaine	312.1588	678.8176	2.89E+08
5	delta9-THC	315.2302	1513.319	6385888
6	Amphetamine	136.0804	419.1944	38130787
7	MA	150.1021	496.8365	3.68E+08
8	MDMA	194.1173	538.9991	6.96E+08
9	Love Drug	180.0866	505.6407	14862760
10	Ketamine	238.0988	599.9313	1.07E+09
11	FM2	314.093	1044.274	2.39E+08
12	Nimetazepam	296.1024	1040.522	2.57E+08

Output file

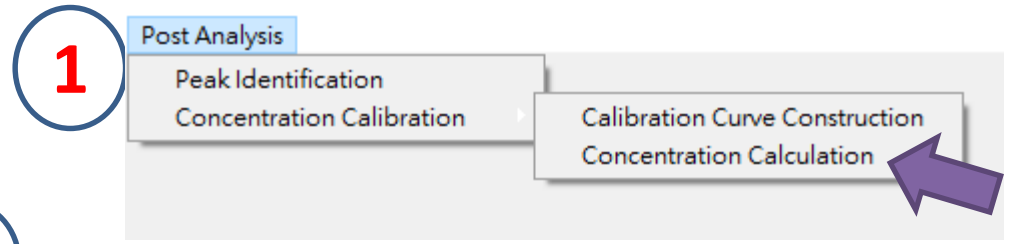
Peak_Inde:	Name	mz	Ret_time.s	T1_500_M
1	Heroin	370.1643	685.3724	507.4039
2	Morphine	286.1431	146.2696	499.5544
3	Cocaine	304.1538	701.136	501.2351
4	Thebaine	312.1588	678.8176	450.4228
5	delta9-THC	315.2302	1513.319	467.1095
6	Amphetam	136.0804	419.1944	773.8217
7	MA	150.1021	496.8365	595.2049
8	MDMA	194.1173	538.9991	520.7164
9	Love Drug	180.0866	505.6407	492.0814
10	Ketamine	238.0988	599.9313	547.7417
11	FM2	314.093	1044.274	512.0763
12	Nimetazepa	296.1024	1040.522	537.2955

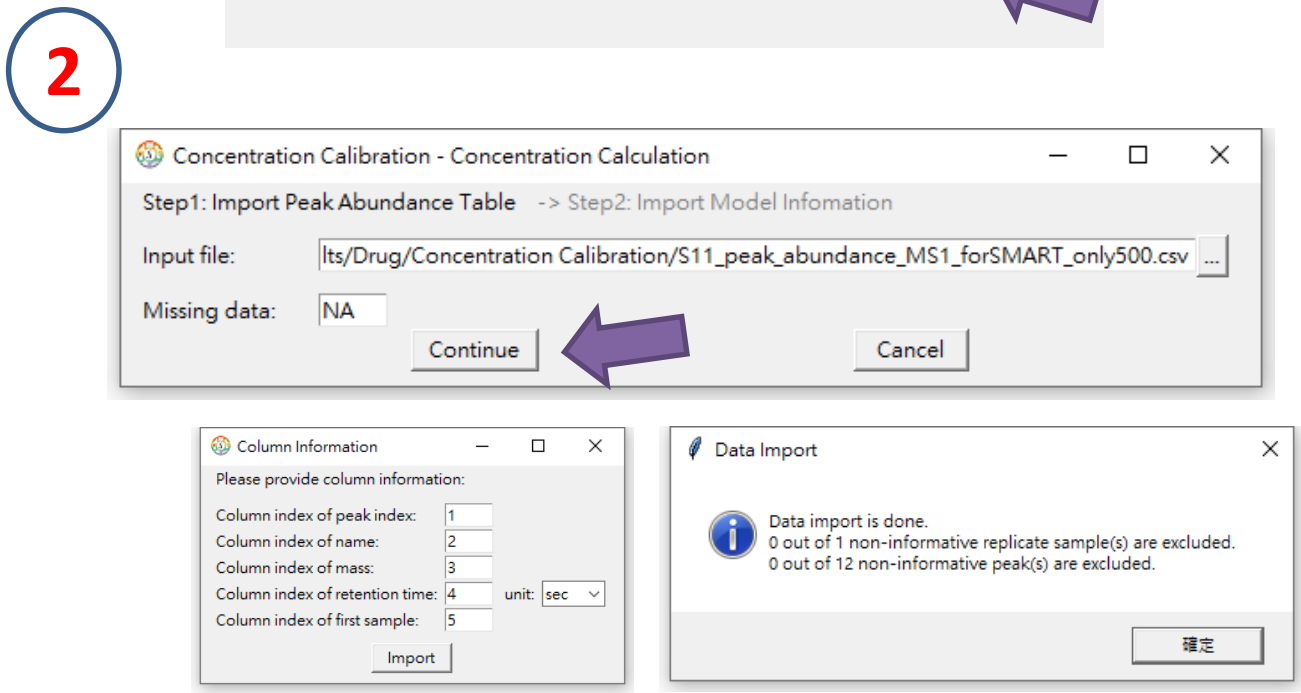


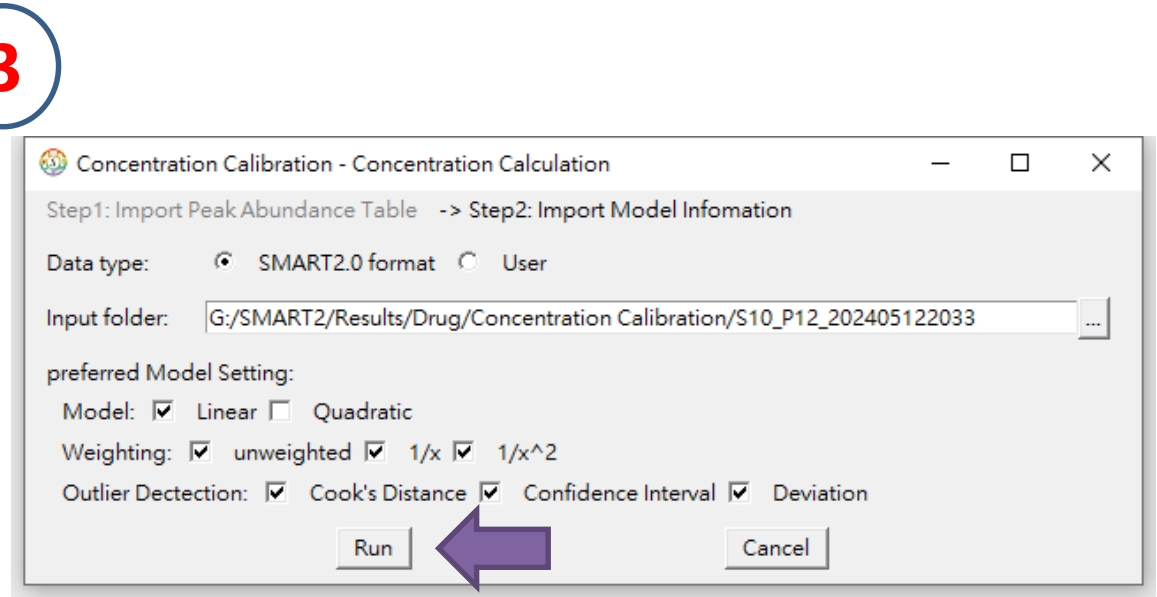
Post analysis (Concentration calibration)

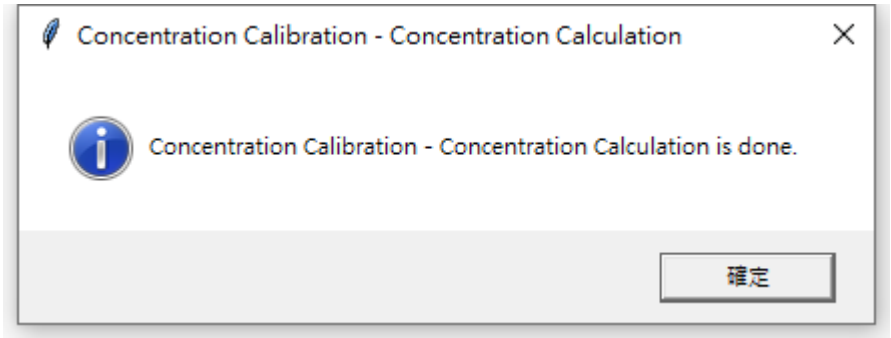
T1_500_MS1

- Concentration calculation
  - Input:
    - Peak abundance table of test samples
    - Model Info. of calibration curve construction









Concentration calculation

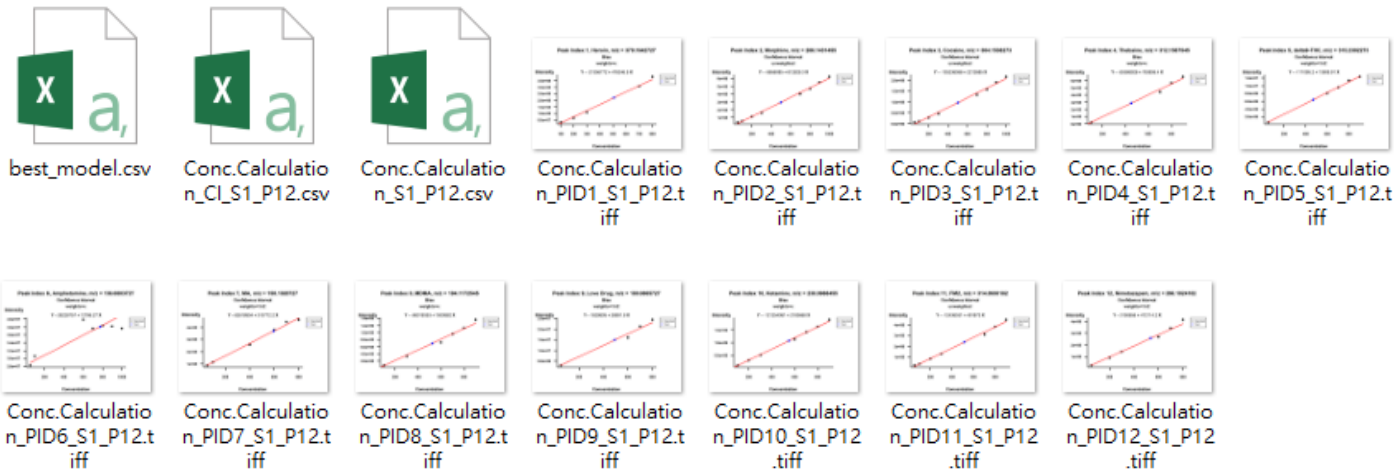
T1_500_MS1

- Output file :
  - Concentration of test samples and figures
    - Conc.Calculation_S1_P12.csv
    - Conc.Calculation_CI_S1_P12.csv
    - best_model.csv

The best model

Peak_Inde	Name	mz	Ret_time.s	model	weights	outlier_det	theta1	theta2
1	Heroin	370.1643	685.3724	linear	weights=x	Bias	-2.1E+07	476246.5
2	Morphine	286.1431	146.2696	linear	unweighte	Confidence	-8656085	612829.3
3	Cocaine	304.1538	701.136	linear	unweighte	Confidence	-1.4E+08	2212685
4	Thebaine	312.1588	678.8176	linear	weights=x	Bias	-5.5E+07	763859.4
5	delta9-THC	315.2302	1513.319	linear	weights=1/	Confidence	-111095	13908.91
6	Amphetam	136.0804	419.1944	linear	weights=x	Confidence	28228757	12796.27
7	MA	150.1021	496.8365	linear	weights=1/	Confidence	62618904	513772.2
8	MDMA	194.1173	538.9991	linear	weights=x	Bias	-8.6E+07	1500882
9	Love Drug	180.0866	505.6407	linear	weights=1/	Bias	1629805	26891.8
10	Ketamine	238.0988	599.9313	linear	weights=x	Bias	-1.2E+08	2183668
11	FM2	314.093	1044.274	linear	unweighte	Confidence	-1.2E+07	491873
12	Nimetazep	296.1024	1040.522	linear	weights=1/	Confidence	3150656	472114.2

Figures



95% CI of Concentration

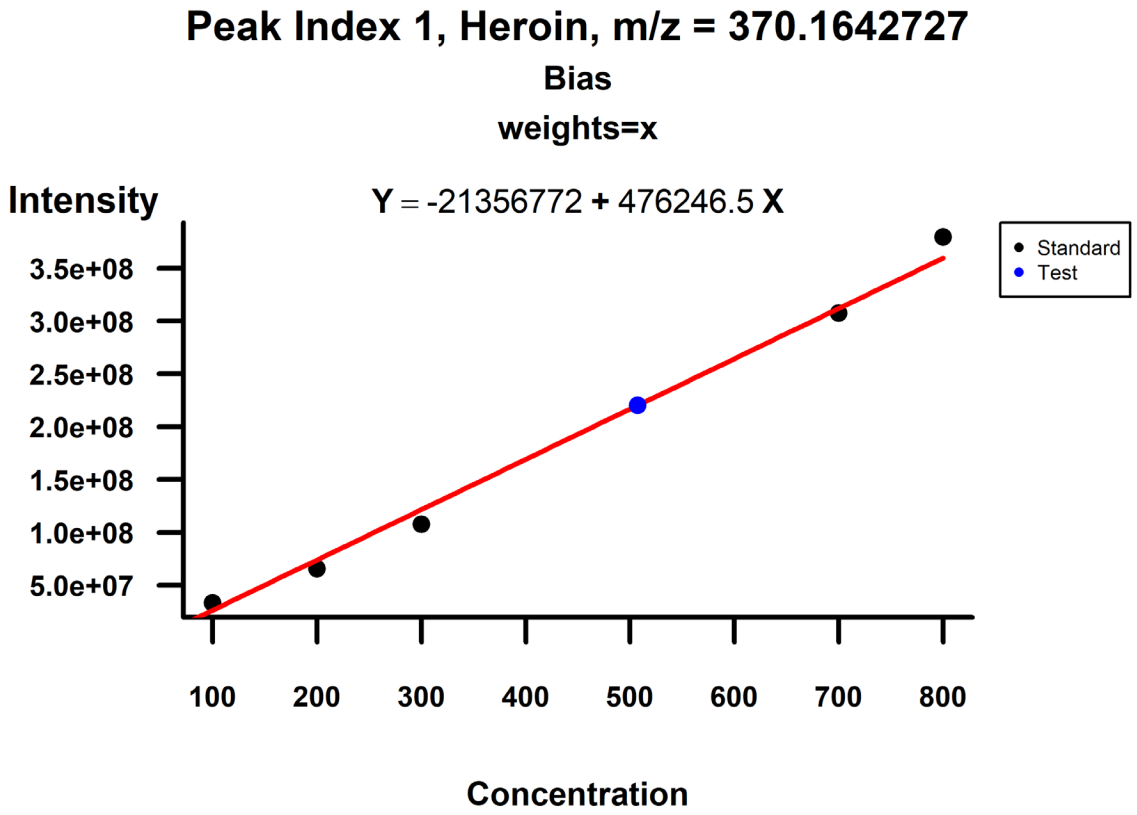
Peak_Inde	Name	mz	Ret_time.s	model	weights	outlier_det	T1_500_M estimate	T1_500_M lower	T1_500_M upper
1	Heroin	370.1643	685.3724	linear	weights=x	Bias	507.4039	453.1017	579.853
2	Morphine	286.1431	146.2696	linear	unweighte	Confidence	499.5544	429.5806	569.4717
3	Cocaine	304.1538	701.136	linear	unweighte	Confidence	501.2351	407.8891	594.506
4	Thebaine	312.1588	678.8176	linear	weights=x	Bias	450.4228	388.4139	524.2525
5	delta9-THC	315.2302	1513.319	linear	weights=1/	Confidence	467.1095	450.9682	484.5766
6	Amphetam	136.0804	419.1944	linear	weights=x	Confidence	773.8217	577.1688	1208.008
7	MA	150.1021	496.8365	linear	weights=1/	Confidence	595.2049	562.0866	632.9944
8	MDMA	194.1173	538.9991	linear	weights=x	Bias	520.7164	477.562	571.1082
9	Love Drug	180.0866	505.6407	linear	weights=1/	Bias	492.0814	401.8802	643.4005
10	Ketamine	238.0988	599.9313	linear	weights=x	Bias	547.7417	530.4445	566.2845
11	FM2	314.093	1044.274	linear	unweighte	Confidence	512.0763	435.9247	589.2633
12	Nimetazep	296.1024	1040.522	linear	weights=1/	Confidence	537.2955	494.561	589.2658

Concentration

Peak_Inde	Name	mz	Ret_time.s	T1_500_M
1	Heroin	370.1643	685.3724	507.4039
2	Morphine	286.1431	146.2696	499.5544
3	Cocaine	304.1538	701.136	501.2351
4	Thebaine	312.1588	678.8176	450.4228
5	delta9-THC	315.2302	1513.319	467.1095
6	Amphetam	136.0804	419.1944	773.8217
7	MA	150.1021	496.8365	595.2049
8	MDMA	194.1173	538.9991	520.7164
9	Love Drug	180.0866	505.6407	492.0814
10	Ketamine	238.0988	599.9313	547.7417
11	FM2	314.093	1044.274	512.0763
12	Nimetazep	296.1024	1040.522	537.2955

Concentration calculation

- Output file :
  - Figures (per drug)
    - Conc.Calculation_PID1_S1_P12.tiff
  - Files
    - Conc.Calculation_S1_P12.csv
    - Conc.Calculation_CI_S1_P12.csv
    - best_model.csv



Concentration calculation

- Output file :
  - Figures (per drug)
    - Conc.Calculation_PID1_S1_P12.tiff
  - Files
    - Conc.Calculation_S1_P12.csv
    - Conc.Calculation_CI_S1_P12.csv
    - best_model.csv

Peak_Index	Name	mz	Ret_time.sec	T1_500_MS1
1	Heroin	370.1642727	685.3723636	507.4039481
2	Morphine	286.1431455	146.2696364	499.5543803
3	Cocaine	304.1538273	701.136	501.2351081
4	Thebaine	312.1587545	678.8176364	450.4228135
5	delta9-THC	315.2302273	1513.319455	467.1095339
6	Amphetamine	136.0803727	419.1943636	773.8216709
7	MA	150.1020727	496.8365455	595.2048664
8	MDMA	194.1172545	538.9990909	520.7163928
9	Love Drug	180.0865727	505.6407273	492.0814322
10	Ketamine	238.0988455	599.9312727	547.7417022
11	FM2	314.0930182	1044.273818	512.0763418
12	Nimetazepam	296.1024182	1040.522182	537.2955304

Concentration calculation

- Output file :
  - Figures (per drug)
    - Conc.Calculation_PID1_S1_P12.tiff
  - Files
    - Conc.Calculation_S1_P12.csv
    - Conc.Calculation_CI_S1_P12.csv
    - best_model.csv

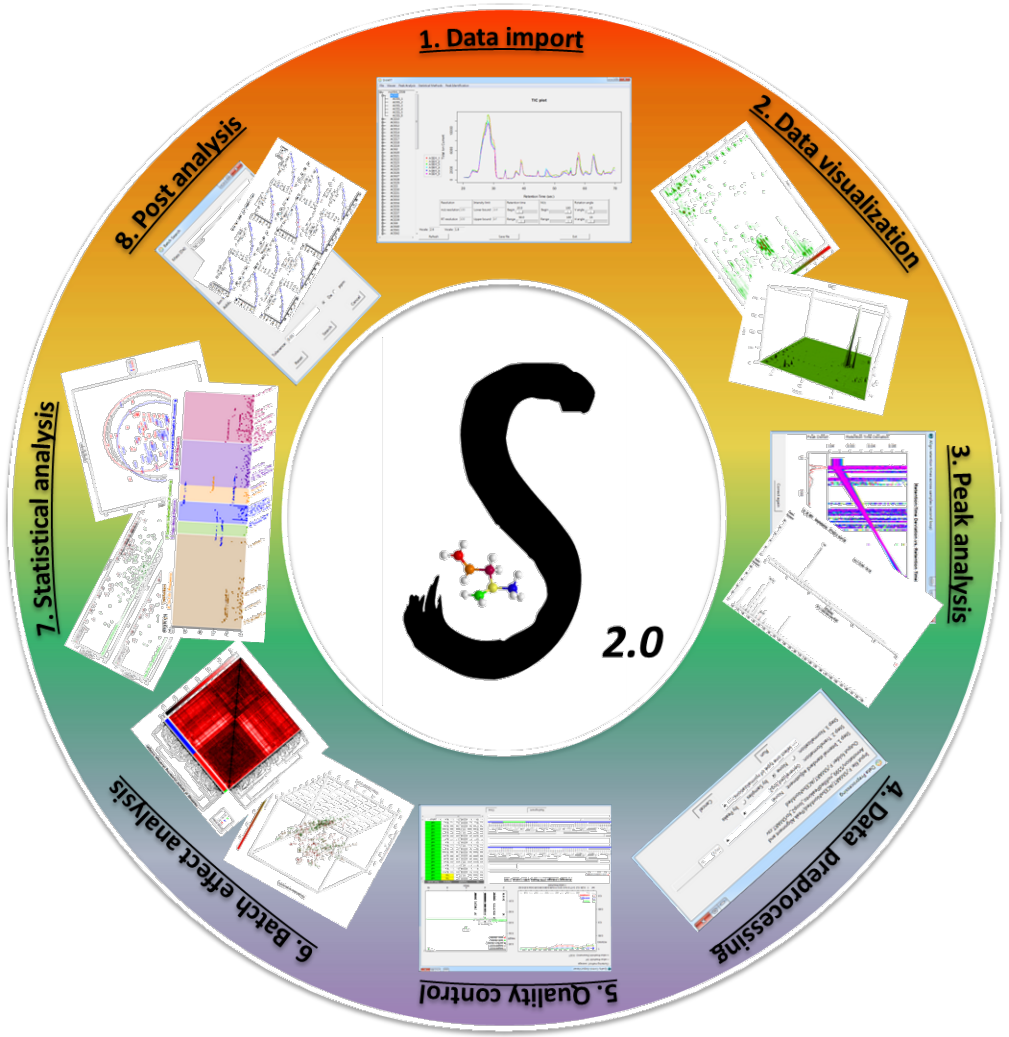
Peak_Index	Name	mz	Ret_time.sec	model	weights	outlier_detection	T1_500_MS1	T1_500_MS1	T1_500_MS1
							estimate	lower	upper
1	Heroin	370.1642727	685.3723636	linear	weights=x	Bias	507.4039481	453.1017284	579.852959
2	Morphine	286.1431455	146.2696364	linear	unweighted	Confidence Interval	499.5543803	429.5806458	569.471698
3	Cocaine	304.1538273	701.136	linear	unweighted	Confidence Interval	501.2351081	407.8890646	594.5059527
4	Thebaine	312.1587545	678.8176364	linear	weights=x	Bias	450.4228135	388.4139437	524.2524615
5	delta9-THC	315.2302273	1513.319455	linear	weights=1/x2	Confidence Interval	467.1095339	450.9681808	484.5765898
6	Amphetamine	136.0803727	419.1943636	linear	weights=x	Confidence Interval	773.8216709	577.1687899	1208.00774
7	MA	150.1020727	496.8365455	linear	weights=1/x2	Confidence Interval	595.2048664	562.0866302	632.9944493
8	MDMA	194.1172545	538.9990909	linear	weights=x	Bias	520.7163928	477.5620114	571.1081788
9	Love Drug	180.0865727	505.6407273	linear	weights=1/x2	Bias	492.0814322	401.8801958	643.4004515
10	Ketamine	238.0988455	599.9312727	linear	weights=x	Bias	547.7417022	530.4445343	566.2844961
11	FM2	314.0930182	1044.273818	linear	unweighted	Confidence Interval	512.0763418	435.9246784	589.2633238
12	Nimetazepam	296.1024182	1040.522182	linear	weights=1/x2	Confidence Interval	537.2955304	494.5609818	589.2657992

Concentration calculation

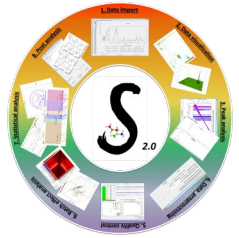
- Output file :
  - Figures (per drug)
    - Conc.Calculation_PID1_S1_P12.tiff
  - Files
    - Conc.Calculation_S1_P12.csv
    - Conc.Calculation_CI_S1_P12.csv
    - best_model.csv

Peak_Index	Name	mz	Ret_time.sec	model	weights	outlier_detection	theta1	theta2
1	Heroin	370.1642727	685.3723636	linear	weights=x	Bias	-21356771.92	476246.4938
2	Morphine	286.1431455	146.2696364	linear	unweighted	Confidence Interval	-8656085.301	612829.2678
3	Cocaine	304.1538273	701.136	linear	unweighted	Confidence Interval	-135236596.4	2212685.355
4	Thebaine	312.1587545	678.8176364	linear	weights=x	Bias	-55368926.41	763859.4158
5	delta9-THC	315.2302273	1513.319455	linear	weights=1/x2	Confidence Interval	-111095.1716	13908.90747
6	Amphetamine	136.0803727	419.1943636	linear	weights=x	Confidence Interval	28228756.57	12796.26888
7	MA	150.1020727	496.8365455	linear	weights=1/x2	Confidence Interval	62618903.73	513772.2091
8	MDMA	194.1172545	538.9990909	linear	weights=x	Bias	-86018352.72	1500881.862
9	Love Drug	180.0865727	505.6407273	linear	weights=1/x2	Bias	1629804.848	26891.79976
10	Ketamine	238.0988455	599.9312727	linear	weights=x	Bias	-121254060.5	2183668.181
11	FM2	314.0930182	1044.273818	linear	unweighted	Confidence Interval	-12436531.12	491872.9913
12	Nimetazepam	296.1024182	1040.522182	linear	weights=1/x2	Confidence Interval	3150656.25	472114.1824

CONCLUSION AND DISCUSSION



Conclusion and Discussion



代謝組學研究面臨的挑戰有很多，以下是其中一些主要的困難：

- **樣本準備和處理:** 樣本的準備和處理對代謝組學的結果影響很大。樣本的收集、保存、提取和前處理都需要非常謹慎，以確保結果的可靠性和可重複性。
- **代謝產物的識別和定量:** 代謝組學研究需要識別和定量大量的代謝產物，這可能需要使用質譜學、核磁共振等高分辨率的分析技術。識別未知的代謝物是一個挑戰，因為代謝物的結構和特性可能非常多樣，且存在許多未知的代謝物。
- **數據處理和分析:** 代謝組學產生的數據量通常非常大，需要複雜的統計和計算方法來處理和分析。這包括校正樣本間的差異、識別生物學上有意義的變化、進行多變量分析等。
- **生物學解釋:** 從代謝組學數據中獲得生物學上有意義的信息是一個挑戰。代謝組學研究通常需要與其他組學數據（如基因組學、蛋白質組學）結合，以更好地理解代謝途徑和生物學過程。
- **樣本異質性:** 生物體內的代謝異質性可能會導致代謝組學研究的難度增加。例如，不同組織、不同個體之間的代謝物水平可能存在差異，這需要在研究設計和分析中加以考慮。

總的來說，代謝組學研究需要多學科的結合，包括化學、生物學、數學和計算機科學等，以應對這些困難並取得有意義的結果。

Other info.

- Database
  - HMDB
    - https://hmdb.ca/
  - MassBank
    - https://massbank.eu/MassBank/
  - Metabolomics Workbench
    - https://www.metabolomicsworkbench.org/
  - MetaboLights
    - https://www.ebi.ac.uk/metabolights
  - LIPID MAPS
    - https://www.lipidmaps.org/
  - METLIN
    - https://metlin.scripps.edu/landing_page.php?pgcontent=mainPage
- Tools
  - MetaboAnalyst 6.0
    - https://www.metaboanalyst.ca/
  - MetDNA2
    - http://metdna.zhulab.cn/
  - MetaCore
    - https://portal.genego.com/
  - Ingenuity Pathway Analysis (IPA)
    - https://digitalinsights.qiagen.com/products-overview/discovery-insights-portfolio/analysis-and-visualization/qiagen-ipa/

E-mail: lyj626@stat.sinica.edu.tw