Basic Use of XCMS -- Local

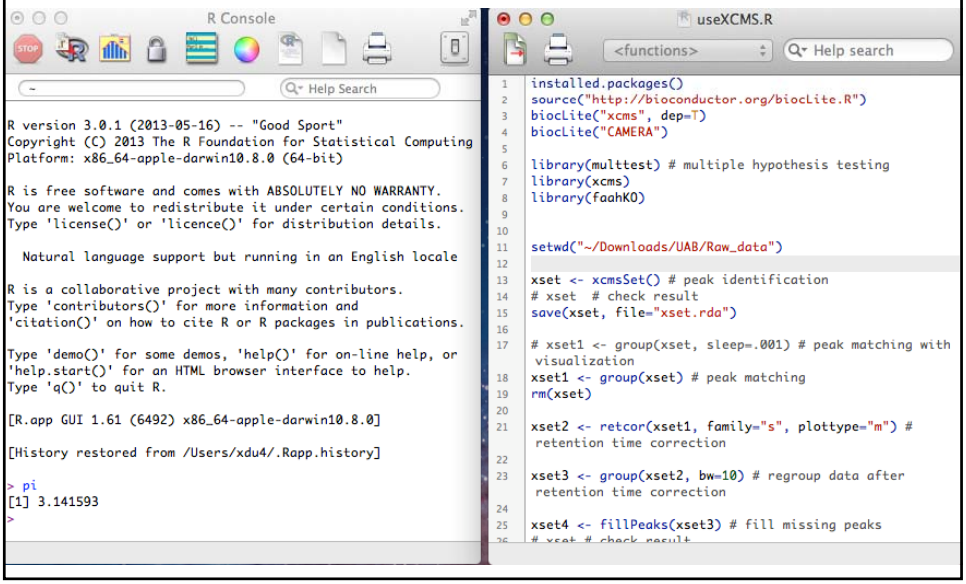
Xiuxia Du
Department of Bioinformatics & Genomics
University of North Carolina at Charlotte

Preparation

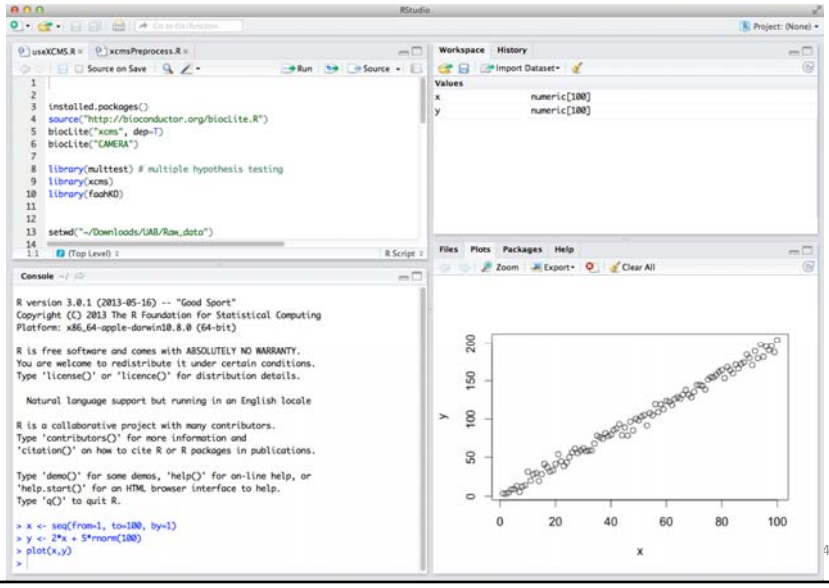
- Required: install R
- Optional: install Rstudio, an IDE (Integrated Development Environment) for R

2

R


R Console
R version 3.0.1 (2013-05-16) -- "Good Sport"
Copyright (C) 2013 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin10.8.0 (64-bit)
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
Natural language support but running in an English locale
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
[R.app GUI 1.61 (6492) x86_64-apple-darwin10.8.0]
[History restored from /Users/xdu4/.Rapp.history]
> pi
[1] 3.141593
useXCMS.R
installed.packages()
source("http://bioconductor.org/biocLite.R")
biocLite("xcms", dep=T)
biocLite("CAMERA")
library(multtest) # multiple hypothesis testing
library(xcms)
library(faahKO)
setwd("~/Downloads/UAB/Raw_data")
xset <- xcmsSet() # peak identification
# xset # check result
save(xset, file="xset.rda")
# xset1 <- group(xset, sleep=.001) # peak matching with visualization
xset1 <- group(xset) # peak matching
rm(xset)
xset2 <- retcor(xset1, family="s", plottype="m") # retention time correction
xset3 <- group(xset2, bw=10) # regroup data after retention time correction
xset4 <- fillPeaks(xset3) # fill missing peaks

RStudio


RStudio
Workspace
History
Values
x numeric[100]
y numeric[100]
Files
Plots
Packages
Help
Console
R version 3.0.1 (2013-05-16) -- "Good Sport"
> x <- seq(from=1, to=100, by=1)
> y <- 2*x + 5*rnorm(100)
> plot(x,y)
y
x
0
50
100
150
200
0
20
40
60
80
100

Get help documents

- Three ways
  - http://www.bioconductor.org/packages/release/bioc/html/xcms.html
  - Google XCMS bioconductor
  - Google XCMS → Scripps Center for Metabolomics and Mass Spectrometry – XCMS → installation → XCMS bioconductor

PDF		FlowChart.pdf
PDF	R Script	Grouping FTICR-MS data with xcms
PDF	R Script	Installation Instructions for xcms
PDF	R Script	LC/MS Preprocessing and Analysis with xcms
PDF	R Script	Processing Tandem-MS and MS^n data with xcms
PDF		Reference Manual
Text		NEWS
Text		LICENSE

Document: step-by-step preprocessing

LC/MS Preprocessing and Analysis with xcms

Colin A. Smith

October 14, 2013

Introduction

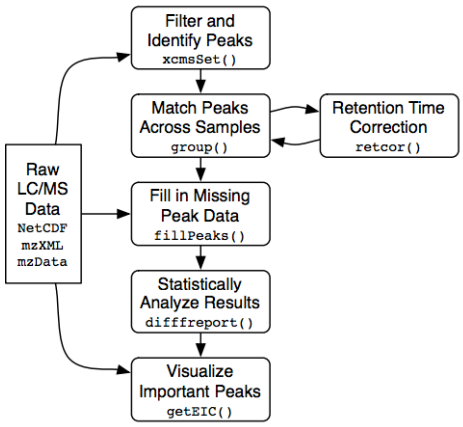
This document describes how to use *xcms* to preprocess LC/MS data for relative quantitation and statistical analysis. It gives examples of how visualization can be used throughout the process and to display final results. An overview of the preprocessing/analysis methodology, along with the function names associated with each step, is shown in Figure 1.

1 Raw Data File Preparation

The *xcms* package reads full-scan LC/MS data from AIA/ANDI format NetCDF, mzXML, and mzData files. All data to be analyzed by *xcms* must first be converted to one of those file formats. Software packages for many instruments are be able to export to NetCDF. For information about how to export to NetCDF, please consult the documentation that

6

XCMS workflow



7

Install and load XCMS packages (I)

Check if the XCMS package has been installed in R

Answer: No

```
> installed.packages()
          Package      LibPath                                  Version
base      "base"       "C:/Program Files/R/R-3.0.2/library" "3.0.2"
boot      "boot"       "C:/Program Files/R/R-3.0.2/library" "1.3-9"
class     "class"      "C:/Program Files/R/R-3.0.2/library" "7.3-9"
cluster   "cluster"    "C:/Program Files/R/R-3.0.2/library" "1.14.4"
codetools "codetools"  "C:/Program Files/R/R-3.0.2/library" "0.2-8"
compiler  "compiler"   "C:/Program Files/R/R-3.0.2/library" "3.0.2"
datasets  "datasets"   "C:/Program Files/R/R-3.0.2/library" "3.0.2"
foreign   "foreign"    "C:/Program Files/R/R-3.0.2/library" "0.8-55"
graphics  "graphics"   "C:/Program Files/R/R-3.0.2/library" "3.0.2"
grDevices "grDevices"  "C:/Program Files/R/R-3.0.2/library" "3.0.2"
grid      "grid"       "C:/Program Files/R/R-3.0.2/library" "3.0.2"
KernSmooth "KernSmooth" "C:/Program Files/R/R-3.0.2/library" "2.23-10"
lattice   "lattice"    "C:/Program Files/R/R-3.0.2/library" "0.20-23"
manipulate "manipulate" "C:/Program Files/R/R-3.0.2/library" "0.98.490"
MASS      "MASS"       "C:/Program Files/R/R-3.0.2/library" "7.3-29"
Matrix    "Matrix"     "C:/Program Files/R/R-3.0.2/library" "1.0-14"
methods   "methods"    "C:/Program Files/R/R-3.0.2/library" "3.0.2"
mgcv      "mgcv"       "C:/Program Files/R/R-3.0.2/library" "1.7-26"
nlme      "nlme"       "C:/Program Files/R/R-3.0.2/library" "3.1-111"
nnet      "nnet"       "C:/Program Files/R/R-3.0.2/library" "7.3-7"
parallel  "parallel"   "C:/Program Files/R/R-3.0.2/library" "3.0.2"
rpart     "rpart"      "C:/Program Files/R/R-3.0.2/library" "4.1-3"
rstudio   "rstudio"    "C:/Program Files/R/R-3.0.2/library" "0.98.490"
spatial   "spatial"    "C:/Program Files/R/R-3.0.2/library" "7.3-7"
splines   "splines"    "C:/Program Files/R/R-3.0.2/library" "3.0.2"
stats     "stats"      "C:/Program Files/R/R-3.0.2/library" "3.0.2"
stats4    "stats4"     "C:/Program Files/R/R-3.0.2/library" "3.0.2"
survival  "survival"   "C:/Program Files/R/R-3.0.2/library" "2.37-4"
tcltk     "tcltk"      "C:/Program Files/R/R-3.0.2/library" "3.0.2"
tools     "tools"      "C:/Program Files/R/R-3.0.2/library" "3.0.2"
utils     "utils"      "C:/Program Files/R/R-3.0.2/library" "3.0.2"
```

Install and load the XCMS packages (II)

```
> source("http://bioconductor.org/biocLite.R")
trying URL 'http://www.bioconductor.org/packages/2.13/bioc/bin/windows/contrib/3.0/BiocInstaller
_1.12.0.zip'
Content type 'application/zip' length 52475 bytes (51 Kb)
opened URL
downloaded 51 Kb

package 'BiocInstaller' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
 C:\Documents and Settings\admin\Local Settings\Temp\RtmpaOkbjd\downloaded_packages
Bioconductor version 2.13 (BiocInstaller 1.12.0), ?biocLite for help
> biocLite("xcms", dep=T)
BioC_mirror: http://bioconductor.org
Using Bioconductor version 2.13 (BiocInstaller 1.12.0), R version 3.0.2.
Installing package(s) 'xcms'
also installing the dependencies 'Rcpp', 'waveslim', 'mzR', 'BiocGenerics', 'Biobase', 'faahKO',
'msdata', 'ncdf', 'multtest', 'rgl', 'MassSpecWavelet', 'RANN', 'RUnit'

trying URL 'http://cran.rstudio.com/bin/windows/contrib/3.0/Rcpp_0.10.6.zip'
Content type 'application/zip' length 3437389 bytes (3.3 Mb)
opened URL
downloaded 3.3 Mb

trying URL 'http://cran.rstudio.com/bin/windows/contrib/3.0/waveslim_1.7.1.zip'
Content type 'application/zip' length 926561 bytes (904 Kb)
opened URL
downloaded 904 Kb
```


Check again.

XCMS and a few other packages have been installed.

```
> installed.packages()
                Package           LibPath                                 Version
base            "base"            "C:/Program Files/R/R-3.0.2/library" "3.0.2"
Biobase         "Biobase"         "C:/Program Files/R/R-3.0.2/library" "2.22.0"
BiocGenerics    "BiocGenerics"    "C:/Program Files/R/R-3.0.2/library" "0.8.0"
BiocInstaller   "BiocInstaller"   "C:/Program Files/R/R-3.0.2/library" "1.12.0"
boot            "boot"            "C:/Program Files/R/R-3.0.2/library" "1.3-9"
class           "class"           "C:/Program Files/R/R-3.0.2/library" "7.3-9"
cluster         "cluster"         "C:/Program Files/R/R-3.0.2/library" "1.14.4"
codetools       "codetools"       "C:/Program Files/R/R-3.0.2/library" "0.2-8"
compiler        "compiler"        "C:/Program Files/R/R-3.0.2/library" "3.0.2"
datasets        "datasets"        "C:/Program Files/R/R-3.0.2/library" "3.0.2"
faahKO          "faahKO"          "C:/Program Files/R/R-3.0.2/library" "1.2.16"
foreign         "foreign"         "C:/Program Files/R/R-3.0.2/library" "0.8-57"
graphics        "graphics"        "C:/Program Files/R/R-3.0.2/library" "3.0.2"
grDevices       "grDevices"       "C:/Program Files/R/R-3.0.2/library" "3.0.2"
grid            "grid"            "C:/Program Files/R/R-3.0.2/library" "3.0.2"
KernSmooth      "KernSmooth"      "C:/Program Files/R/R-3.0.2/library" "2.23-10"
lattice         "lattice"         "C:/Program Files/R/R-3.0.2/library" "0.20-24"
manipulate      "manipulate"      "C:/Program Files/R/R-3.0.2/library" "0.98.490"
MASS            "MASS"            "C:/Program Files/R/R-3.0.2/library" "7.3-29"
MassSpecWavelet "MassSpecWavelet" "C:/Program Files/R/R-3.0.2/library" "1.28.0"
Matrix          "Matrix"          "C:/Program Files/R/R-3.0.2/library" "1.1-0"
methods         "methods"         "C:/Program Files/R/R-3.0.2/library" "3.0.2"
mgcv            "mgcv"            "C:/Program Files/R/R-3.0.2/library" "1.7-27"
msdata          "msdata"          "C:/Program Files/R/R-3.0.2/library" "0.1.15"
multtest        "multtest"        "C:/Program Files/R/R-3.0.2/library" "2.18.0"
mzR             "mzR"             "C:/Program Files/R/R-3.0.2/library" "1.8.0"
ncdf            "ncdf"            "C:/Program Files/R/R-3.0.2/library" "1.6.6"
nlme            "nlme"            "C:/Program Files/R/R-3.0.2/library" "3.1-113"
nnet            "nnet"            "C:/Program Files/R/R-3.0.2/library" "7.3-7"
parallel        "parallel"        "C:/Program Files/R/R-3.0.2/library" "3.0.2"
RANN            "RANN"            "C:/Program Files/R/R-3.0.2/library" "2.3.0"
Rcpp            "Rcpp"            "C:/Program Files/R/R-3.0.2/library" "0.10.6"
rgl             "rgl"             "C:/Program Files/R/R-3.0.2/library" "0.93.991"
rpart           "rpart"           "C:/Program Files/R/R-3.0.2/library" "4.1-4"
rstudio         "rstudio"         "C:/Program Files/R/R-3.0.2/library" "0.98.490"
RUnit           "RUnit"           "C:/Program Files/R/R-3.0.2/library" "0.4.26"
spatial         "spatial"         "C:/Program Files/R/R-3.0.2/library" "7.3-7"
splines         "splines"         "C:/Program Files/R/R-3.0.2/library" "3.0.2"
stats           "stats"           "C:/Program Files/R/R-3.0.2/library" "3.0.2"
stats4          "stats4"          "C:/Program Files/R/R-3.0.2/library" "3.0.2"
survival        "survival"        "C:/Program Files/R/R-3.0.2/library" "2.37-4"
tcltk           "tcltk"           "C:/Program Files/R/R-3.0.2/library" "3.0.2"
tools           "tools"           "C:/Program Files/R/R-3.0.2/library" "3.0.2"
utils           "utils"           "C:/Program Files/R/R-3.0.2/library" "3.0.2"
waveslim        "waveslim"        "C:/Program Files/R/R-3.0.2/library" "1.7.1"
xcms            "xcms"            "C:/Program Files/R/R-3.0.2/library" "1.38.0"
```

Install and load the XCMS packages (IV)

```
> library(multtest)
> library(xcms)
> library(faahKO)
>
```

← for multiple hypothesis testing

← demo data supplied by XCMS

11

Raw data format and organization

- Open formats that XCMS can read
  - AIA/ANDI NetCDF
  - mzData
  - mzXML
- Organization
  - Use sub-directories that correspond to sample class information
- Datasets for demonstration
  - faahKO data package supplied by XCMS
  - Data is stored in netCDF format.
  - The raw data sets are stored in a folder on your computer.

12

Raw data preparation (I)

In R command window on Mac:

```
> cdfpath <- system.file("cdf", package = "faahKO")
> list.files(cdfpath, recursive = TRUE)
 [1] "KO/ko15.CDF" "KO/ko16.CDF" "KO/ko18.CDF" "KO/ko19.CDF" "KO/ko21.CDF"
 [6] "KO/ko22.CDF" "WT/wt15.CDF" "WT/wt16.CDF" "WT/wt18.CDF" "WT/wt19.CDF"
[11] "WT/wt21.CDF" "WT/wt22.CDF"
> cdfpath
[1] "/Library/Frameworks/R.framework/Versions/3.0/Resources/library/faahKO/cdf"
>
```

From the terminal, check where the datasets are on your computer:

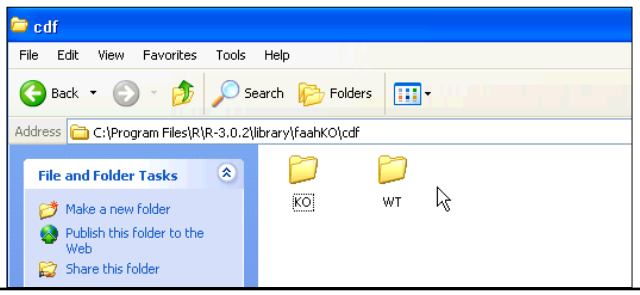
```
dhcp00056:WT xdu4$ cd /Library/Frameworks/R.framework/Versions/3.0/Resources/library/faahKO/cdf
dhcp00056:cdf xdu4$ ls -lp
total 0
drwxr-xr-x  8 xdu4  admin  272 Aug  7 18:00 KO/
drwxr-xr-x  8 xdu4  admin  272 Aug  7 18:00 WT/
dhcp00056:cdf xdu4$ cd KO
dhcp00056:KO xdu4$ ls
ko15.CDF        ko16.CDF        ko18.CDF        ko19.CDF        ko21.CDF        ko22.CDF
dhcp00056:KO xdu4$ cd ..
dhcp00056:cdf xdu4$ cd WT
dhcp00056:WT xdu4$ ls
wt15.CDF        wt16.CDF        wt18.CDF        wt19.CDF        wt21.CDF        wt22.CDF
dhcp00056:WT xdu4$
```

Raw data preparation (II)

In R command window on Windows:

```
> cdfpath <- system.file("cdf", package = "faahKO")
> list.files(cdfpath, recursive = TRUE)
 [1] "KO/ko15.CDF" "KO/ko16.CDF" "KO/ko18.CDF" "KO/ko19.CDF" "KO/ko21.CDF" "KO/ko22.CDF"
 [7] "WT/wt15.CDF" "WT/wt16.CDF" "WT/wt18.CDF" "WT/wt19.CDF" "WT/wt21.CDF" "WT/wt22.CDF"
> cdfpath
[1] "C:/Program Files/R/R-3.0.2/library/faahKO/cdf"
>
```

Check where the datasets are on your computer:



14

Raw data preparation (III)

Get the list of the raw data files:

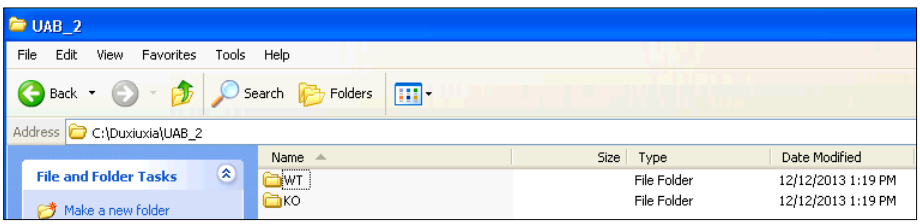
```
> cdffiles <- list.files(cdfpath, recursive = TRUE, full.names = TRUE)
> cdffiles
 [1] "/Library/Frameworks/R.framework/Versions/3.0/Resources/library/faahKO/cdf/KO/ko15.CDF"
 [2] "/Library/Frameworks/R.framework/Versions/3.0/Resources/library/faahKO/cdf/KO/ko16.CDF"
 [3] "/Library/Frameworks/R.framework/Versions/3.0/Resources/library/faahKO/cdf/KO/ko18.CDF"
 [4] "/Library/Frameworks/R.framework/Versions/3.0/Resources/library/faahKO/cdf/KO/ko19.CDF"
 [5] "/Library/Frameworks/R.framework/Versions/3.0/Resources/library/faahKO/cdf/KO/ko21.CDF"
 [6] "/Library/Frameworks/R.framework/Versions/3.0/Resources/library/faahKO/cdf/KO/ko22.CDF"
 [7] "/Library/Frameworks/R.framework/Versions/3.0/Resources/library/faahKO/cdf/WT/wt15.CDF"
 [8] "/Library/Frameworks/R.framework/Versions/3.0/Resources/library/faahKO/cdf/WT/wt16.CDF"
 [9] "/Library/Frameworks/R.framework/Versions/3.0/Resources/library/faahKO/cdf/WT/wt18.CDF"
[10] "/Library/Frameworks/R.framework/Versions/3.0/Resources/library/faahKO/cdf/WT/wt19.CDF"
[11] "/Library/Frameworks/R.framework/Versions/3.0/Resources/library/faahKO/cdf/WT/wt21.CDF"
[12] "/Library/Frameworks/R.framework/Versions/3.0/Resources/library/faahKO/cdf/WT/wt22.CDF"
>
```

15

Raw data preparation (IV)

- Alternatively
  - Specify the working directory
  - By default, XCMS will recursively search through the current working directory for NetCDF/mzXML/mzData files.

```
> my_working_directory <- paste("C:", "Duxiuxia", "UAB_2", sep=.Platform$file.sep) # on Windows
> setwd(my_working_directory)
> my_working_directory
[1] "C:/Duxiuxia/UAB_2"
```



16

Peak identification (I)

- Command: `xcmsSet()`

```
> xset <- xcmsSet(cdffiles)
250:38 300:103 350:226 400:338 450:431 500:529 550:674 600:847
250:43 300:128 350:275 400:394 450:500 500:637 550:835 600:1027
250:25 300:93 350:227 400:337 450:411 500:498 550:640 600:758
250:19 300:67 350:169 400:258 450:301 500:373 550:488 600:580
250:24 300:60 350:166 400:254 450:315 500:391 550:501 600:582
250:31 300:71 350:183 400:280 450:338 500:422 550:532 600:604
250:41 300:105 350:212 400:319 450:416 500:533 550:684 600:838
250:27 300:107 350:232 400:347 450:440 500:549 550:712 600:905
250:24 300:87 350:200 400:293 450:351 500:426 550:548 600:661
250:22 300:65 350:161 400:243 450:293 500:358 550:483 600:561
250:28 300:69 350:157 400:229 450:282 500:364 550:493 600:592
250:30 300:81 350:188 400:280 450:356 500:473 550:618 600:765
>
```

- One separate row for a dataset
- For each pair of numbers, the first number is the m/z XCMS is currently processing. The second number is the number of peaks that have been identified so far.

17

Peak identification (II)

- If raw data files are in your working directory, then:

```
> xset <- xcmsSet()
250:38 300:103 350:226 400:338 450:431 500:529 550:674 600:847
250:43 300:128 350:275 400:394 450:500 500:637 550:835 600:1027
250:25 300:93 350:227 400:337 450:411 500:498 550:640 600:758
250:19 300:67 350:169 400:258 450:301 500:373 550:488 600:580
250:24 300:60 350:166 400:254 450:315 500:391 550:501 600:582
250:31 300:71 350:183 400:280 450:338 500:422 550:532 600:604
250:41 300:105 350:212 400:319 450:416 500:533 550:684 600:838
250:27 300:107 350:232 400:347 450:440 500:549 550:712 600:905
250:24 300:87 350:200 400:293 450:351 500:426 550:548 600:661
250:22 300:65 350:161 400:243 450:293 500:358 550:483 600:561
250:28 300:69 350:157 400:229 450:282 500:364 550:493 600:592
250:30 300:81 350:188 400:280 450:356 500:473 550:618 600:765
>
```

18

Peak Identificaiton (III)

- Take a look at the *xcmsSet* object:

```
> xset
An "xcmsSet" object with 12 samples

Time range: 2506.1-4147.7 seconds (41.8-69.1 minutes)
Mass range: 200.1-599.3338 m/z
Peaks: 4721 (about 393 per sample)
Peak Groups: 0
Sample classes: KO, WT

Profile settings: method = bin
                  step = 0.1

Memory usage: 0.713 MB
>
```

19

Peak identification (IV)

- The default parameters should work acceptably in most cases.
  - Default peak detection method: *matched filter*
  - Alternative approach: *centWave* for high resolution MS data
- However, a number of parameters might need to be optimized for particular instruments or experimental conditions.
  - Matched filtration: model peak width, m/z step size for creating extracted ion base peak chromatograms (EIBPC), the algorithm to create EIBPC, …
  - centWave: ppm, peak width range, …
- To be explained in the next section "Parameter set-up …" by Paul

20

Matching peaks across samples

- After peak identification, peaks representing the same analyte across samples must be placed into groups.
- This is accomplished with the `group()` method.

```
> xset <- group(xset)
262 325 387 450 512 575
>
```

- There are several grouping parameters to consider optimizing for your chromatography and mass spectrometer (to be explained by Paul).

21

Retention time correction (I)

- XCMS uses peak groups to identify and correct drifts in retention time from run to run.
- Only well-behaved peak groups are used: missing the peak from at most one sample and having at most one extra peak.
- These parameters can be changed with the *missing* and *extra* arguments.
- For each of those well-behaved groups, XCMS calculates a median retention time and, for every sample, a deviation from that median.

22

Retention time correction (II)

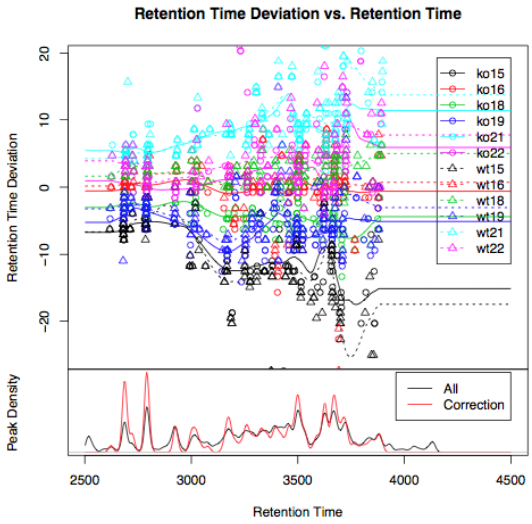
- Within a sample, the observed deviation generally changes over time in a nonlinear fashion.
- Those changes are modeled using a local polynomial regression technique.
- Retention time correction is performed by the `retcor()` method.

```
> xset2 <- retcor(xset, family = "symmetric", plottype = "mdevden")
Retention Time Correction Groups: 133
>
```

- The *plottype* argument produces the plot on the next slide.

23

Retention time correction (III)



24

Retention time correction (IV)

- Use the plot to supervise the algorithm. The plot includes data points used for regression and the resulting deviation profiles.
- The plot also shows the distribution of peak groups across retention time.

25

Retention time correction (V)

- After retention time correction, the initial peak grouping becomes invalid.
- Peak re-grouping is needed.

```
> xset2 <- group(xset2, bw = 10)
262 325 387 450 512 575
>
```

- This iteration of peak grouping and alignment can be repeated in an iterative fashion.

26

Filling in missing peaks (I)

- Peaks could be missing due to imperfection in peak identification or because an analyte was not present in a sample.
- For missing peaks that correspond to analytes that are actually in the sample, the missing data points can be filled in by re-reading the raw data files and integrating them in the regions of the missing peaks.
- This is performed using the `fillPeaks()` method.

27

Filling in missing peaks (II)

```
> xset3 <- fillPeaks(xset2)
/Library/Frameworks/R.framework/Versions/3.0/Resources/library/faahKO/cdf/KO/ko15.CDF
/Library/Frameworks/R.framework/Versions/3.0/Resources/library/faahKO/cdf/KO/ko16.CDF
/Library/Frameworks/R.framework/Versions/3.0/Resources/library/faahKO/cdf/KO/ko18.CDF
/Library/Frameworks/R.framework/Versions/3.0/Resources/library/faahKO/cdf/KO/ko19.CDF
/Library/Frameworks/R.framework/Versions/3.0/Resources/library/faahKO/cdf/KO/ko21.CDF
/Library/Frameworks/R.framework/Versions/3.0/Resources/library/faahKO/cdf/KO/ko22.CDF
/Library/Frameworks/R.framework/Versions/3.0/Resources/library/faahKO/cdf/WT/wt15.CDF
/Library/Frameworks/R.framework/Versions/3.0/Resources/library/faahKO/cdf/WT/wt16.CDF
/Library/Frameworks/R.framework/Versions/3.0/Resources/library/faahKO/cdf/WT/wt18.CDF
/Library/Frameworks/R.framework/Versions/3.0/Resources/library/faahKO/cdf/WT/wt19.CDF
/Library/Frameworks/R.framework/Versions/3.0/Resources/library/faahKO/cdf/WT/wt21.CDF
/Library/Frameworks/R.framework/Versions/3.0/Resources/library/faahKO/cdf/WT/wt22.CDF
```

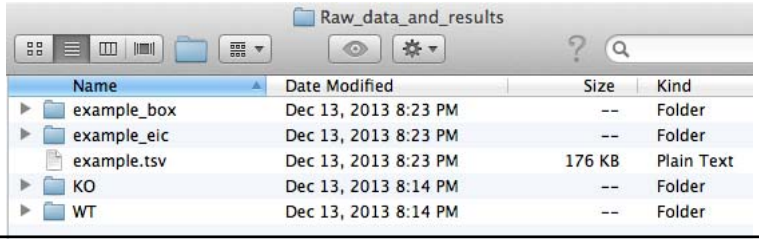
28

Analyzing and visualizing results (I)

- A report showing the most statistically significant differences in analyte intensities can be generated with the `diffreport()` method.

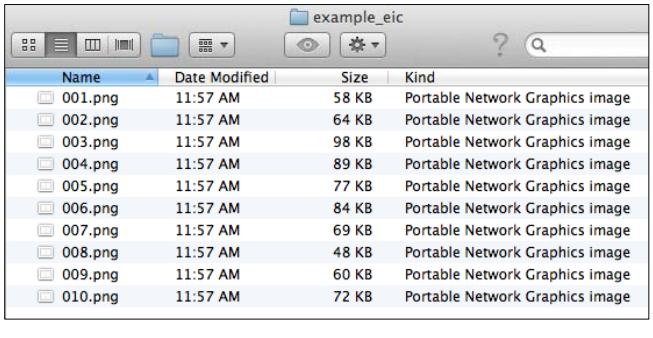
```
> reporttab <- diffreport(xset3, "WT", "KO", "example", 10, metlin = 0.15, h=480, w=640)
ko15 ko16 ko18 ko19 ko21 ko22 wt15 wt16 wt18 wt19 wt21 wt22
>
```

- Results are stored in two folders and one spread sheet file.

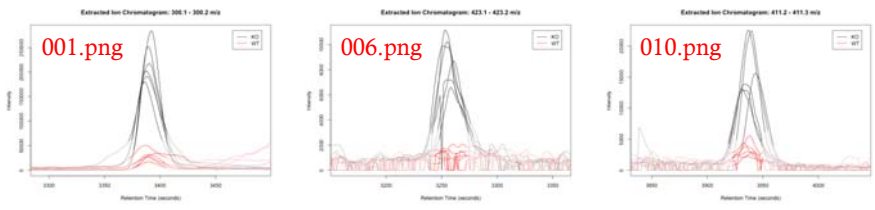


29

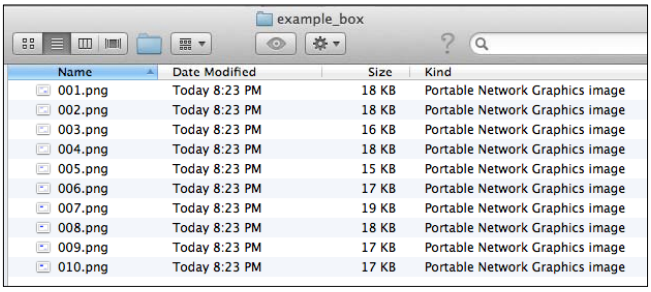
Analyzing and visualizing results (II)



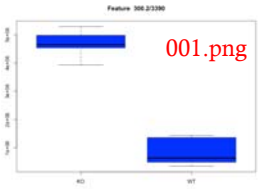
Extracted ion chromatograms for significant ions

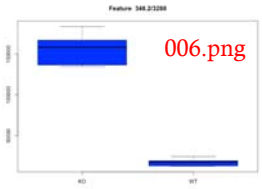


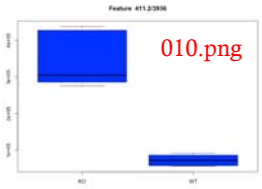
Analyzing and visualizing results (III)



Box plots for significant ions







Analyzing and visualizing results (IV)

example.tsv (column A to K)

	A	B	C	D	E	F	G	H	I	J	K
		name	fold	tstat	pvalue	mzmed	mzmin	mzmax	rtmed	rtmin	rtmax
1		M300T3390	5.69359438	14.4436804	5.03E-08	300.189812	300.170629	300.200012	3390.32362	3386.76499	3396.33492
2		M301T3390	5.8765876	15.5756963	6.71E-08	301.187865	301.16586	301.194921	3389.62744	3386.76499	3392.1013
3		M298T3187	3.8709182	11.9389063	3.31E-07	298.150848	298.105429	298.159194	3186.80296	3184.12371	3191.31223
4		M491T3397	24.975703	16.8398648	4.46E-06	491.200012	491.187713	491.20635	3397.16034	3367.12282	3424.68101
5		M348T3288	9.00506717	17.2626531	5.03E-06	348.162043	348.123398	348.174149	3288.24803	3284.66356	3294.1279
6		M423T3257	6.24617282	10.8189245	4.71E-05	423.149887	423.106429	423.157354	3256.61095	3254.85456	3261.53255
7		M327T3419	26.7484435	11.0506059	9.44E-05	327.198887	327.168262	327.200012	3419.47081	3412.91699	3427.2301
8		M326T3417	15.5444373	10.4238641	0.00013344	326.200012	326.170474	326.200012	3417.01426	3411.30668	3425.61712
9		M410T3938	6.78734161	9.11467023	0.00016577	410.26577	410.212496	410.29861	3937.70017	3932.86967	3946.13523
10		M411T3936	4.63296841	8.7449582	0.0002105	411.23392	411.224299	411.286168	3936.49194	3932.86967	3945.26839
11		M322T3390	7.11832905	8.34304436	0.00026398	322.121163	322.100006	322.162268	3390.49164	3389.62744	3392.1013
12		M449T3291	5.68627603	7.9070938	0.00032076	449.132096	449.100006	449.167376	3291.36493	3289.63734	3295.49023
13		M354T3618	8.38720426	7.43940447	0.00055487	354.200012	354.200012	354.300018	3618.08658	3616.78057	3620.93729
14		M330T3632	10.5951954	6.68193606	0.00097609	330.190995	330.141883	330.214943	3631.5667	3627.77112	3635.01114
15		M348T3419	12.6099923	6.30073621	0.00142646	348.143096	348.109896	348.18278	3419.47081	3409.77379	3428.84201
16		M328T3633	10.2967672	6.12160764	0.00164693	328.200012	328.193341	328.200012	3633.3197	3625.18912	3635.88794
17		M329T3631	10.3980966	5.92879556	0.0018321	329.200012	329.197015	329.200012	3631.33547	3626.61376	3635.77785
18		M357T3836	4.65856087	5.04631827	0.00228991	357.206233	357.197998	357.219202	3835.93832	3833.86436	3837.21789

32

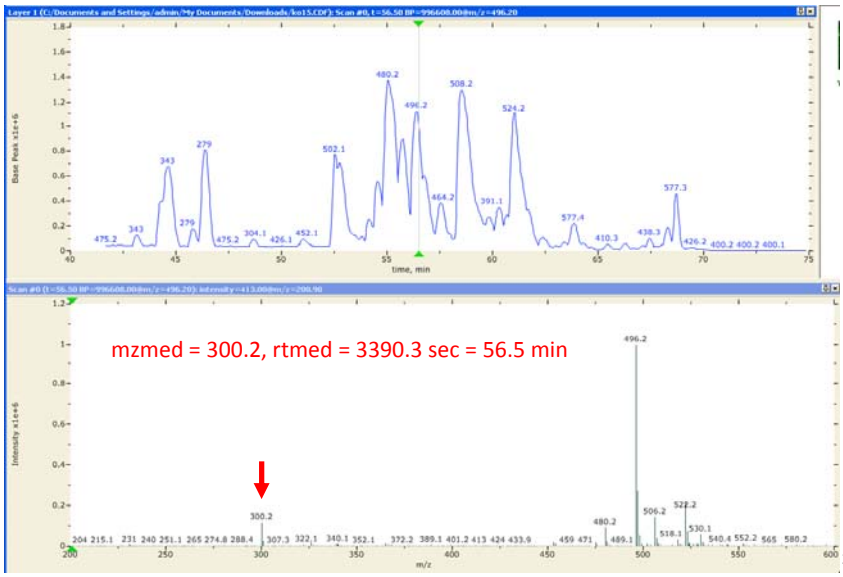
Analyzing and visualizing results (V)

example.tsv (column O to AA)

O	P	Q	R	S	T	U	V	W	X	Y	Z	AA
metlin	ko15	ko16	ko18	ko19	ko21	ko22	wt15	wt16	wt18	wt19	wt21	wt22
http://metlir	4534353.62	4980914.48	5290739.14	4564262.9	4733236.08	3931592.59	349660.885	491793.181	645526.705	634108.849	1438254.45	1364627.84
http://metlir	962353.43	1047934.14	1109303.04	946943.393	984787.205	806171.473	86450.4116	120096.52	143007.948	137319.686	218483.143	291392.971
http://metlir	180780.817	203926.952	191015.911	190626.85	156869.08	220288.622	16269.096	43677.784	54739.1289	76318.0077	54726.1154	49679.9425
http://metlir	432037.001	332159.073	386966.751	334951.453	294816.236	373577.608	7643.13807	10519.9422	26472.293	33598.3228	8030.46734	0
http://metlir	165830.903	183665.01	150844.995	134637.117	136452.456	167008.053	24302.8057	16631.3863	19213.6611	12822.5391	12452.1286	18789.754
http://metlir	236249.547	255168.616	212710.776	180690.967	191746.741	152861.186	29530.1749	17037.2075	35133.1225	27706.2092	47228.7374	40193.5296
http://metlir	1108851.28	950126.529	674222.714	677091.267	772290.305	1013977.91	58898.7043	21990.98	27642.9741	31727.8261	45299.168	8715.57767
http://metlir	4809521.45	3931304.64	2913711.55	2819100.93	3284987.28	4346409.7	259229.273	314154.323	161900.584	200939.74	247976.865	237853.621
http://metlir	980012.03	1319875.61	1375014.96	949318.774	899985.582	799401.791	137450.635	203674.39	223544.907	160796.141	95447.3273	110763.407
http://metlir	304741.632	426745.765	437071.769	304811.059	285571.596	275335.275	57976.3758	92450.81	86924.8392	79420.9306	56605.4873	65708.7501
http://metlir	238208.464	272654.011	378948.167	273593.949	247266.87	179812.074	23463.0482	30685.2109	62726.3335	42656.2854	32328.4535	31575.6239
http://metlir	144480.636	149070.729	95314.6844	136823.134	83447.365	163903.723	15781.0387	14619.5489	19411.8176	23885.3028	33471.3336	28779.3836
http://metlir	1135610.95	904522.365	832780.32	532867.709	570306.904	729939.966	112167.159	154475.666	51277.225	91243.2653	79838.1181	72094.7396
http://metlir	201588.695	167042.482	143015.845	105447.719	92547.8575	95737.3812	18621.1758	18071.6701	3745.92163	17251.6947	9713.90424	8609.33088
http://metlir	456678.246	324908.659	190365.213	192504.5	328211.717	439481.04	38080.0815	27120.0761	17746.3781	17365.864	21200.2003	31711.0758
http://metlir	6463382.87	5630552.38	4398403.51	3085430.38	3289887.97	2626186.41	402409.001	573880.805	403748.596	412858.155	385802.429	297208.544
http://metlir	1521141.36	1315446.89	990625.769	707798.716	759301.22	605854.743	114347.574	140213.418	111356.479	69186.0779	88438.2624	43885.9531
http://metlir	328565.947	567776.99	391437.155	254805.159	305032.229	218120.799	54773.6586	148383.91	77678.498	66290.1789	26578.2065	69723.9671

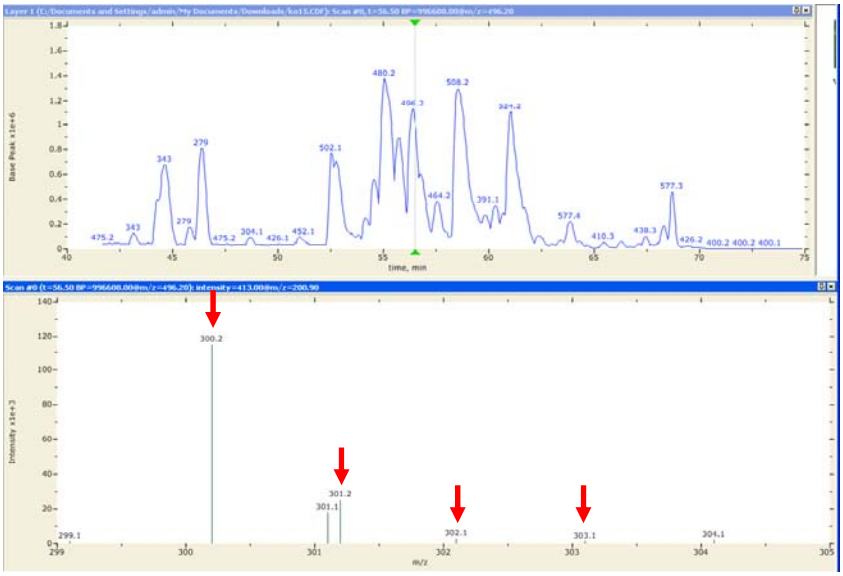
33

Going back to the raw data (I)



34

Going back to the raw data (II)



Visualize raw data

- Mass spectrometer vendors save data in proprietary formats.
- These data files can be converted to open data formats for easy reading.
- Software tools to do the conversion: msConvert

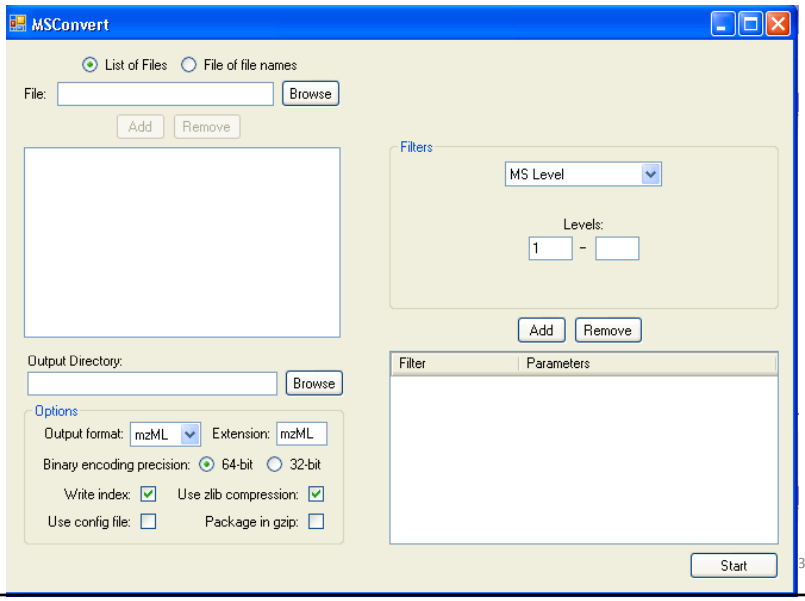
36

msConvert (I)

- Part of ProteoWizard
- Read from
  - mzML, mzXML, MGF
  - Agilent, Bruker, Thermo, Waters, ABSciex
- Write to
  - open formats
  - perform various filters and transformations
- http://proteowizard.sourceforge.net/
- For Windows, msConvertGUI is available for easy file conversion.

37

msConvert (II)



38

Raw data visualization

- Software tool: Insilicos Viewer
  - View raw MS data in formats including mzXML, mzData, mzML, and ANDI CDF
  - http://insilicos.com/products/insilicos-viewer-1
  - Quick demo

39

Run all the commands

useXCMS.R

Source on Save

```
1  # use XCMS for processing LC-MSmetabolomics data
2
3  installed.packages()
4  source("http://bioconductor.org/biocLite.R")
5  biocLite("xcms", dep=T)
6  biocLite("CAMERA")
7
8  library(multtest) # load the package for multiple hypothesis testing
9  library(xcms) # load the XCMS package
10 library(faahKO) # load the dataset supplied by XCMS
11
12 setwd("~/Downloads/UAB_2/Results")
13
14 cdfpath <- system.file("cdf", package = "faahKO")
15 list.files(cdfpath, recursive = TRUE)
16
17 cdffiles <- list.files(cdfpath, recursive = TRUE, full.names = TRUE)
18 xset <- xcmsSet(cdffiles)
```

```
> source("/Users/xdu4/Downloads/UAB_2/Code/useXCMS.R")
```

40

Thank you!

41