

Hands-on Practice: LC-MS Data Pre-processing and Statistical Analysis

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

Outline

- XCMS
- MarkerView™
- MetaboAnalyst

XCMS

- Two ways to use XCMS
 - XCMS Online
 - XCMS in R
- How to get XCMS?
 - <http://metlin.scripps.edu/xcms/>
- XCMS Online

XCMS in R: peak picking

```
> setwd("~/Downloads/UAB/Row_data")
> xset <- xcmsSet()
135:520 185:1343 235:1905 285:2411 335:2874 385:3365 435:3806 485:4295 535:4821 585:5357 635:5828 685:6165 735:6554
785:6981 835:7270 885:7474 935:7628 985:7749
135:488 185:1203 235:1716 285:2211 335:2660 385:3151 435:3585 485:4058 535:4600 585:5116 635:5564 685:5905 735:6291
785:6692 835:6966 885:7163 935:7296 985:7407
135:544 185:1361 235:1921 285:2392 335:2838 385:3326 435:3761 485:4206 535:4743 585:5295 635:5742 685:6084 735:6494
785:6899 835:7188 885:7406 935:7561 985:7671
135:531 185:1316 235:1866 285:2344 335:2780 385:3259 435:3679 485:4113 535:4655 585:5201 635:5634 685:5968 735:6363
785:6772 835:7064 885:7271 935:7436 985:7565
135:644 185:1458 235:1969 285:2344 335:2769 385:3307 435:3677 485:4026 535:4472 585:4847 635:5232 685:5568 735:5945
785:6406 835:6721 885:7015 935:7275 985:7545
135:630 185:1512 235:2065 285:2444 335:2891 385:3463 435:3865 485:4257 535:4754 585:5135 635:5549 685:5911 735:6311
785:6781 835:7124 885:7482 935:7770 985:8084
135:628 185:1500 235:2035 285:2412 335:2860 385:3433 435:3838 485:4235 535:4700 585:5100 635:5500 685:5846 735:6246
785:6722 835:7035 885:7357 935:7615 985:7903
135:689 185:1664 235:2221 285:2613 335:3099 385:3671 435:4084 485:4499 535:4978 585:5376 635:5790 685:6162 735:6538
785:6999 835:7331 885:7649 935:7888 985:8165
>
```

- A separate line for each sample to report the status of processing
- 135:520 = m/z that is currently being processed and number of peaks that has been identified so far

XCMS in R

- Peak picking using `centWave`

XCMS in R: peak grouping

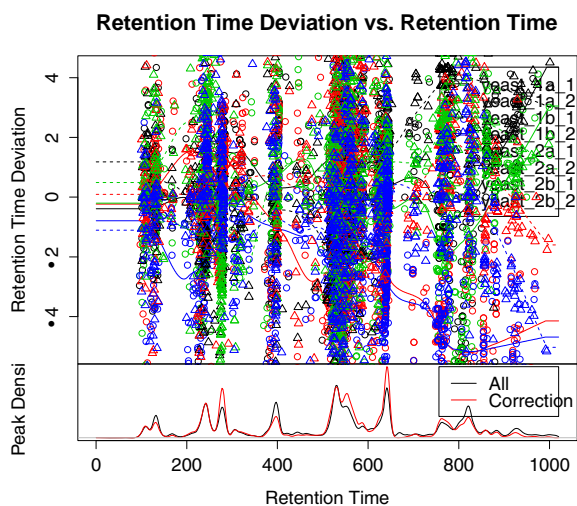
- Peak picking is time **consuming**. Save peak picking results.
- Peaks representing the same analyte across samples are placed into groups.

```
> save(xset, file="xset.rda")  
> xset1 <- group(xset)  
147 210 272 335 397 460 522 585 647 710 772 835 897 960
```

- The algorithm processes the peak lists in order of increasing mass and regularly outputs the mass it is currently working on.

XCMS in R: alignment

```
> xset2 <- retcor(xset1, family="s", plottype="m")
Retention Time Correction Groups: 1313
>
```



XCMS in R

- Re-grouping

```
> xset3 <- group(xset2, bw=10)
147 210 272 335 397 460 522 585 647 710 772 835 897 960
> |
```

- Filling in missing peaks

```
> xset4 <- fillPeaks(xset3)
/Users/xdu4/Downloads/UAB/Raw_data/Group1/yeast_1a_1.mzXML
/Users/xdu4/Downloads/UAB/Raw_data/Group1/yeast_1a_2.mzXML
/Users/xdu4/Downloads/UAB/Raw_data/Group1/yeast_1b_1.mzXML
/Users/xdu4/Downloads/UAB/Raw_data/Group1/yeast_1b_2.mzXML
/Users/xdu4/Downloads/UAB/Raw_data/Group2/yeast_2a_1.mzXML
/Users/xdu4/Downloads/UAB/Raw_data/Group2/yeast_2a_2.mzXML
/Users/xdu4/Downloads/UAB/Raw_data/Group2/yeast_2b_1.mzXML
/Users/xdu4/Downloads/UAB/Raw_data/Group2/yeast_2b_2.mzXML
```

XCMS in R

- Re-grouping

```
> xset3 <- group(xset2, bw=10)
147 210 272 335 397 460 522 585 647 710 772 835 897 960
> |
```

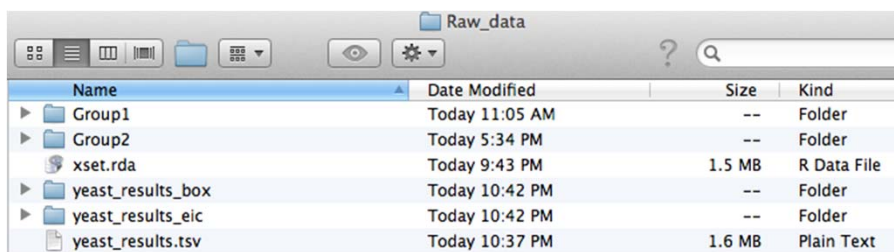
- Filling in missing peaks

```
> xset4 <- fillPeaks(xset3)
/Users/xdu4/Downloads/UAB/Raw_data/Group1/yeast_1a_1.mzXML
/Users/xdu4/Downloads/UAB/Raw_data/Group1/yeast_1a_2.mzXML
/Users/xdu4/Downloads/UAB/Raw_data/Group1/yeast_1b_1.mzXML
/Users/xdu4/Downloads/UAB/Raw_data/Group1/yeast_1b_2.mzXML
/Users/xdu4/Downloads/UAB/Raw_data/Group2/yeast_2a_1.mzXML
/Users/xdu4/Downloads/UAB/Raw_data/Group2/yeast_2a_2.mzXML
/Users/xdu4/Downloads/UAB/Raw_data/Group2/yeast_2b_1.mzXML
/Users/xdu4/Downloads/UAB/Raw_data/Group2/yeast_2b_2.mzXML
```

XCMS in R

- Producing report and visualizing results

```
> reporttab <- diffreport(xset4, "Group1", "Group2", "yeast_results", 10, metlin=0.15)
Loading required package: multtest
yeast_1a_1 yeast_1a_2 yeast_1b_1 yeast_1b_2 yeast_2a_1 yeast_2a_2 yeast_2b_1 yeast_2b_2
> |
```



| Name | Date Modified | Size | Kind |
|---------------------|----------------|--------|-------------|
| ▶ Group1 | Today 11:05 AM | -- | Folder |
| ▶ Group2 | Today 5:34 PM | -- | Folder |
| xset.rda | Today 9:43 PM | 1.5 MB | R Data File |
| ▶ yeast_results_box | Today 10:42 PM | -- | Folder |
| ▶ yeast_results_eic | Today 10:42 PM | -- | Folder |
| yeast_results.tsv | Today 10:37 PM | 1.6 MB | Plain Text |

XCMS in R

| P | Q | R | S | T | U | V | W |
|------------|------------|------------|------------|------------|------------|------------|------------|
| yeast_1a_1 | yeast_1a_2 | yeast_1b_1 | yeast_1b_2 | yeast_2a_1 | yeast_2a_2 | yeast_2b_1 | yeast_2b_2 |
| 34717.4482 | 33358.2004 | 34366.1522 | 34160.3854 | 2882.47275 | 3371.7625 | 2393.51281 | 3300.75382 |
| 26939.3955 | 27779.7 | 27445.856 | 27984.21 | 3669.95545 | 3722.04977 | 3749.32576 | 4671.71583 |
| 102207.809 | 105005.148 | 100611.066 | 100679.99 | 7785.78308 | 5136.79707 | 4846.2887 | 5139.278 |
| 63165.6077 | 62144.6075 | 60761.177 | 62660.6185 | 5787.4851 | 3506.32053 | 3403.52954 | 5955.81549 |
| 26772.69 | 27094.2447 | 27287.685 | 26104.6262 | 4858.24567 | 4841.56112 | 4161.09386 | 4968.6426 |
| 2819.97745 | 2584.71209 | 2225.97814 | 2505.81442 | 11264.052 | 11546.002 | 11065.9882 | 11375.8985 |
| 6018.19125 | 6110.6784 | 6045.88437 | 6376.38914 | 632.749826 | 598.926233 | 666.364121 | 398.625202 |
| 6631.3597 | 6740.60111 | 6998.4285 | 7054.67954 | 892.621763 | 1346.40913 | 1051.28492 | 1115.8443 |
| 14547.7419 | 14402.6092 | 13932.2917 | 14804.0554 | 3307.02624 | 3029.59988 | 2663.02394 | 3135.21593 |
| 20810.8723 | 20612.6993 | 22135.5049 | 21502.5273 | 917.795039 | 1669.91378 | 1907.18266 | 2476.94114 |
| 23634.0423 | 23090.6494 | 22212.0987 | 22398.3907 | 449.803111 | 513.419861 | 301.261968 | 1216.6323 |
| 14975.9615 | 15079.7005 | 14343.7446 | 15416.8042 | 3276.1988 | 3089.91225 | 3393.0719 | 4120.77329 |
| 3366653.38 | 3318335.09 | 3410270.66 | 3532591.59 | 1056876.73 | 1255308.14 | 1090760.74 | 1155758.82 |
| 926781.043 | 980182.659 | 918498.386 | 977798.948 | 132724.68 | 191478.353 | 134536.652 | 159818.885 |
| 41831.2563 | 41717.4947 | 40954.8792 | 41957.5693 | 11672.034 | 12498.8301 | 10842.258 | 10661.3771 |
| 103208.438 | 108192.064 | 98365.0294 | 102729.187 | 8642.03533 | 13657.1007 | 11954.8323 | 19014.6224 |
| 157416.091 | 158541.778 | 159913.37 | 164486.319 | 57494.3354 | 67186.5599 | 58801.17 | 61948.1686 |
| 9033.60652 | 9436.59157 | 8782.25173 | 9186.47879 | 2108.13731 | 2289.81657 | 2290.28708 | 2859.34717 |
| 2084.17779 | 1990.55365 | 2243.65565 | 1853.38918 | 5251.73523 | 5431.07712 | 5195.70669 | 5478.63405 |
| 7065.22523 | 7375.41813 | 7620.31146 | 7159.2192 | 2307.44718 | 2502.27336 | 2636.64809 | 2815.74139 |
| 667964.686 | 684033.236 | 677799.947 | 713397.568 | 213821.873 | 268870.595 | 230433.834 | 243020.007 |
| 3989.373 | 3723.43962 | 4064.56333 | 3815.63479 | 843.760725 | 1052.56913 | 828.776033 | 660.681584 |
| 7256.87931 | 6625.52162 | 7210.95111 | 6902.24477 | 834.21705 | 1015.46051 | 1111.49493 | 1347.16684 |
| 25905.4006 | 24301.1196 | 25780.5831 | 26272.8005 | 8330.34415 | 10303.6976 | 9393.89706 | 9728.02935 |
| 180707.592 | 194703.968 | 192951.112 | 186273.298 | 52480.0109 | 53778.6097 | 61544.5422 | 52430.1668 |
| 17904.523 | 18939.6297 | 17324.7737 | 18364.4337 | 5727.68724 | 5314.96908 | 5242.57422 | 6689.85351 |
| 29527.6099 | 28262.5891 | 28917.7467 | 27223.4385 | 6647.29621 | 5589.6799 | 7386.89971 | 8420.23345 |

Calculate the fold change and t statistic in Excel.

Functions in Excel:

- average()
- var.s()
- sqrt()
- t.test()

t-test in R

- t.test()

XCMS on Bioconductor

- <http://www.bioconductor.org/packages/release/bioc/html/xcms.html>

MarkerView™

- Pre-processing
 - XCMS
 - MarkerView™
- Statistical analysis
 - MeV

Outline

- Pre-processing
 - XCMS
 - MarkerView™
- Statistical analysis
 - MeV

Outline

- Pre-processing
 - XCMS
 - MarkerView™
- Statistical analysis
 - MeV

Outline

- Pre-processing
 - XCMS
 - MarkerView™
- Statistical analysis
 - MeV

Outline

- Pre-processing
 - XCMS
 - MarkerView™
- Statistical analysis
 - MeV