XCMS Workshop: Peak Grouping & LOESS Alignment

May 31, 2009 – ASMS – Philadelphia, PA

Colin A. Smith

Overview

- Kernel density peak grouping

- xcmsSet structure

- LOESS retention time alignment

Generating an xcmsSet

```
> cdfpath <- system.file("cdf",
  package = "faahKO")
> cdffiles <- list.files(cdfpath,
  recursive = TRUE, full.names = TRUE)
> xset <- xcmsSet(cdffiles)
```


Kernel Density Grouping Algorithm

- Detects groups of peaks by finding regions of high peak density
- Uses kernel density estimation
  - Step 1: Generate a 512 bin histogram
  - Step 2: Smooth using a gaussian function of a user-defined width
- Group borders are defined by starting at peaks and rolling downhill
- Parameters:
  - mzwid (0.25 m/z): maximum expected variation in m/z in a group
  - bw (30 seconds): standard deviation of smoothing function
  - minfrac (0.5)/minsamp (1): minimum fraction/number of samples in a given peak group to be considered valid
  - max (50): maximum number of peaks per m/z slice
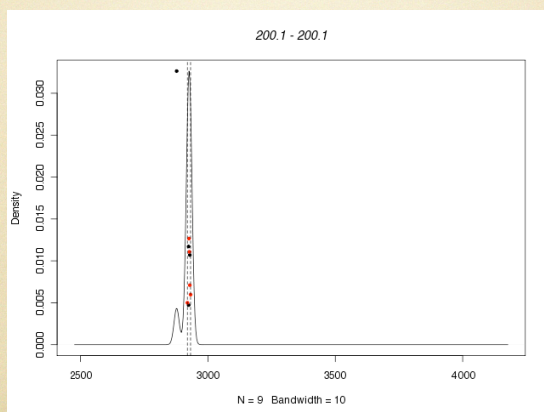
Grouping Plot

```
> png("GroupingMovie%03d.png",
    width=640, height=480, type="cairo1")
> xset <- group(xset, sleep=.00001)
262 325 387 450 512 575
> dev.off()
```

Grouping Plot
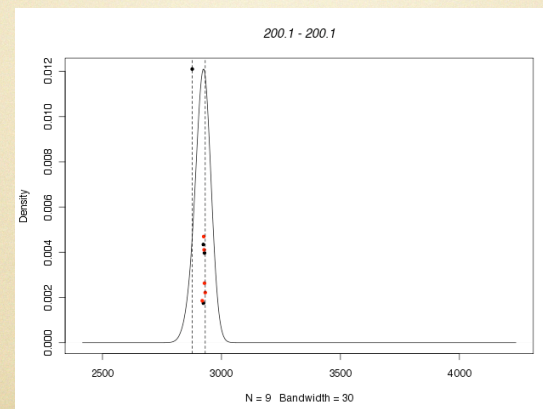
Grouping Plot

Loading Raw Data

```
> 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: 403
Sample classes: KO, WT

Profile settings: method = bin
                  step = 0.1

Memory usage: 0.762 MB
```

Inside xcmsSet

```
> str(xset)
Formal class 'xcmsSet' [package "xcms"] with 9 slots
  ..@ peaks    : num [1:4721, 1:13] 200 201 205 206 207 ...
  .. ..- attr(*, "dimnames")=List of 2
  .. .. ..$ : NULL
  .. .. ..$ : chr [1:13] "mz" "mzmin" "mzmax" "rt" ...
  ..@ groups   : num [1:403, 1:9] 200 205 206 207 219 ...
  ..- attr(*, "dimnames")=List of 2
  .. .. ..$ : NULL
  .. .. ..$ : chr [1:9] "mzmed" "mzmin" "mzmax" "rtmed" ...
  ..@ groupidx :List of 403
  .. ..$ : int [1:9] 1 471 1027 2078 2405 2865 3707 4004 4315
...
  .. ..$ : int [1:5] 1025 1443 1757 3344 4314
  ..@ filled   : int(0)
...
```

Inside xcmsSet

```
  ..@ phenoData:'data.frame':  12 obs. of  1 variable:
  .. ..$ class: Factor w/ 2 levels "KO","WT": 1 1 1 1 1 1 2 2 2
   2 ...
  ..@ rt        :List of 2
  .. ..$ raw      :List of 12
  .. .. ..$ : num [1:1278] 2501 2503 2505 2506 2508 ...
...
  .. .. ..$ : num [1:1278] 2501 2503 2505 2506 2508 ...
  .. ..$ corrected:List of 12
  .. .. ..$ : num [1:1278] 2501 2503 2505 2506 2508 ...
...
  .. .. ..$ : num [1:1278] 2501 2503 2505 2506 2508 ...
  ..@ filepaths: chr [1:12] "/Users/Colin/Library/R/librar..." ...
  ..@ profinfo :List of 2
  .. ..$ method: chr "bin"
  .. ..$ step  : num 0.1
  ..@ polarity : chr(0)
```

groups Matrix

```
> head(xset@groups)
        mzmed    mzmin    mzmax    rtmed    rtmin    rtmax npeaks KO WT
[1,] 200.1000 200.1000 200.1000 2925.480 2876.967 2931.740      9  4  5
[2,] 205.0000 205.0000 205.0000 2790.894 2784.635 2795.591     12  6  6
[3,] 205.9927 205.9786 206.0023 2790.112 2784.635 2795.591     12  6  6
[4,] 207.0850 207.0440 207.1000 2718.906 2712.647 2726.731     12  6  6
[5,] 219.0848 219.0488 219.1000 2524.852 2518.592 2529.547      9  4  5
[6,] 231.0236 231.0000 231.0812 2517.029 2509.202 2535.807      6  3  3
```

groups Matrix

LOESS Retention Time Alignmet

- Find "well behaved" groups defined as having few missing peaks and few extra peaks

- For each group, calculate the median and determine deviations from the median for every sample

- For every sample, use a local regression method called LOESS to determine a nonlinear deviation profile

- Parameters:

  - missing (1)/extra (1): maximums for peak groups

  - method ("loess"): deviation profile fitting method

  - span (0.2): fraction of data to use for local fitting

  - family ("gaussian"): LOESS fitting method, "symmetric" allows outlier removal

Median

Retention Time
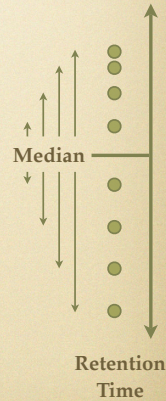
```
> xset2 <- retcor(xset, span=0.2) # 0.2 default
```
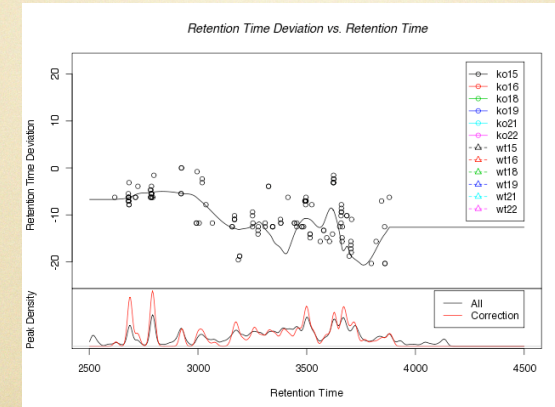
```
> xset2 <- retcor(xset, span=0.5)
```
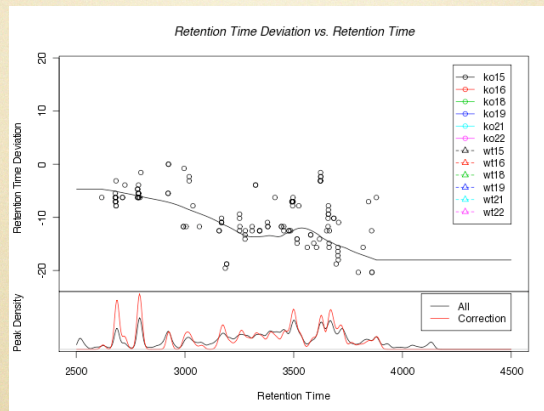
```
> xset2 <- retcor(xset, span=0.05)
```

```
> xset2 <- retcor(xset, method="linear")
```
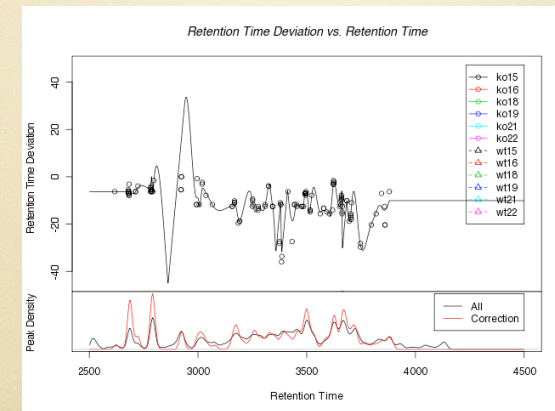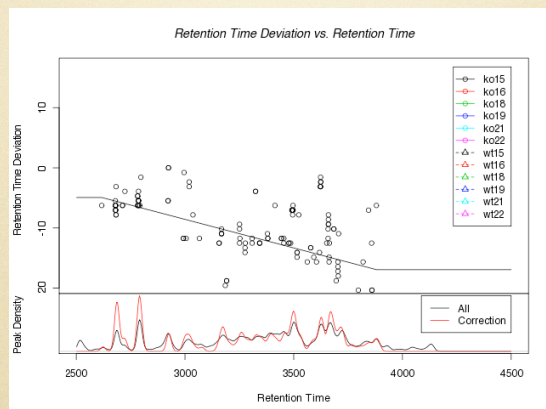
Extracting a Matrix of Grouped Peak Data

```
> head(groupval(xset))
```

	ko15	ko16	ko18	ko19	ko21	ko22	wt15	wt16	wt18	wt19	wt21	wt22
200.1/2925	1	471	1027	NA	NA	2078	2405	2865	NA	3707	4004	4315
205/2791	3	473	1029	1445	1758	2079	2407	2866	3349	3708	4005	4316
206/2790	4	474	1030	1446	1759	2080	2408	2867	3350	3709	4006	4317
207.1/2719	5	475	1031	1447	1760	2081	2409	2868	3351	3710	4007	4318
219.1/2525	8	477	NA	NA	1761	2084	2413	2871	3352	NA	4008	4319
231/2517	NA	NA	1032	1448	1762	NA	NA	NA	NA	3711	4009	4320

Extracting a Matrix of Grouped Peak Data

```
> head(groupval(xset, value="into"))
```

	ko15	ko16	ko18	ko19	ko21	ko22
200.1/2925	147887.5	451600.7	65290.38	NA	NA	162012.4
205/2791	1778568.9	1567038.1	1482796.38	1039129.8	1223132.4	1072037.7
206/2790	237993.6	269714.0	201393.42	150107.3	176989.7	156797.0
207.1/2719	380873.0	460629.7	351750.14	219288.0	286848.6	235022.6
219.1/2525	235544.9	173623.4	NA	NA	185792.4	174458.8
231/2517	NA	NA	222609.07	286232.1	435094.5	NA

	wt15	wt16	wt18	wt19	wt21	wt22
200.1/2925	175177.1	82619.48	NA	69198.22	153273.5	98144.28
205/2791	1950287.5	1466780.60	1572679.16	1275312.76	1356014.3	1231442.16
206/2790	276541.8	222366.15	211717.71	186850.88	188285.9	172348.76
207.1/2719	417169.6	324892.46	277990.70	220972.35	252874.0	236728.16
219.1/2525	244584.5	161184.05	72029.38	NA	238194.4	173829.95
231/2517	NA	NA	NA	240261.21	201316.2	179437.72