Supplementary Information for Joint bounding of peaks across samples improves differential analysis in mass spectrometry-based metabolomics

Leslie Myint[1], Andre Kleensang[2], Liang Zhao[2], Thomas Hartung[2,3], and Kasper D. Hansen[1,4,*]

[1]Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, 21205 MD, Baltimore [2]Center for Alternatives to Animal Testing (CAAT), Department of Environmental Health and Engineering, Johns Hopkins Bloomberg School of Public Health, 21205 Baltimore, MD, USA [3]University of Konstanz, CAAT-Europe, Germany

[4]McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, 21205
Baltimore, MD, USA

Content:

1. Supplementary methods describing XCMS and MZmine2 parameters used to process example data.
2. Supplementary Figures S1-S13.

*To whom correspondence should be addressed. Email: khansen@jhsph.edu

Supplementary Methods

XCMS parameters

Data available in centroid mode was processed with the `xcmsSet` command with the "centWave" method. Profile mode data was processed with the "matchedFilter" method. Parameter optimization was performed with the IPO package availabe on Bioconductor. Optimization was performed on a subset of two samples for each dataset. These two samples were chosen to be the one in each sample group that had the largest total intensity (integrated area beneath the total ion chromatogram) because these were expected to have the richest set of peak information. In most cases we used the default starting parameters for the optimizations (obtained with the `getDefaultXcmsSetStartingParams` and `getDefaultRetGroupStartingParams` functions). The signal-to-noise threshold value is not optimized by default, but we optimized it by setting the starting parameters with

```
snthresh = c(3,8)
```

We also optimized the prefilter values with the following starting parameters

```
prefilter = c(2,3)
prefilter_value = c(200,300)
```

We used the default starting parameters for retention time alignment and grouping optimization with the exception of the MTBLS2 and MTBLS213 datasets.

Due to optimization running times in excess of 11 days, we modified the starting parameters for the two MTBLS2 datasets as follows to match the parameters given in the original paper.

```
min_peakwidth = c(5,12)
max_peakwidth = c(12,35)
prefilter = 3
prefilter_value = 200
snthresh = 5
ppm = 25
minfrac = 0.75
```

For the same reason, we modified the starting parameters for the MTBLS213 dataset as follows to match the parameters given in the original paper.

```
min_peakwidth = c(5,12)
max_peakwidth = c(20,35)
ppm = 30
minfrac = 0.5
```

After obtaining optimized parameters, we ran the `xcmsSet` command followed by `group` using the "density" method, `retcor` with the "obiwarp" method, and finally `fillPeaks`.

MZmine2 parameters

We ran MZmine2 version 2.21 with the GridMass - 2D peak detection procedure, the join aligner for retention time alignment and grouping, and the same-range gap filler module. To the best of our knowledge, there is no automated method for obtaining optimized MZmine parameters, so we translated optimized XCMS parameters to MZmine parameters as follows.

GridMass peak detection

- Minimum height: use optimized prefilter value from XCMS
- M/Z Tolerance: use optimized mzdiff from XCMS unless negative. If negative, use 100*optimized XCMS ppm/1e6.
- Min-max width time (in minutes): use optimized minimum and maximum peak width from XCMS multiplied by 60 to convert to minutes
- Smoothing M/Z: use 0.5*M/Z tolerance as this parameters is recommended to be smaller than the m/z tolerance
- Intensity similarity ratio: the default 0.5 was used
- Ignore times: the default of no times ignored was used

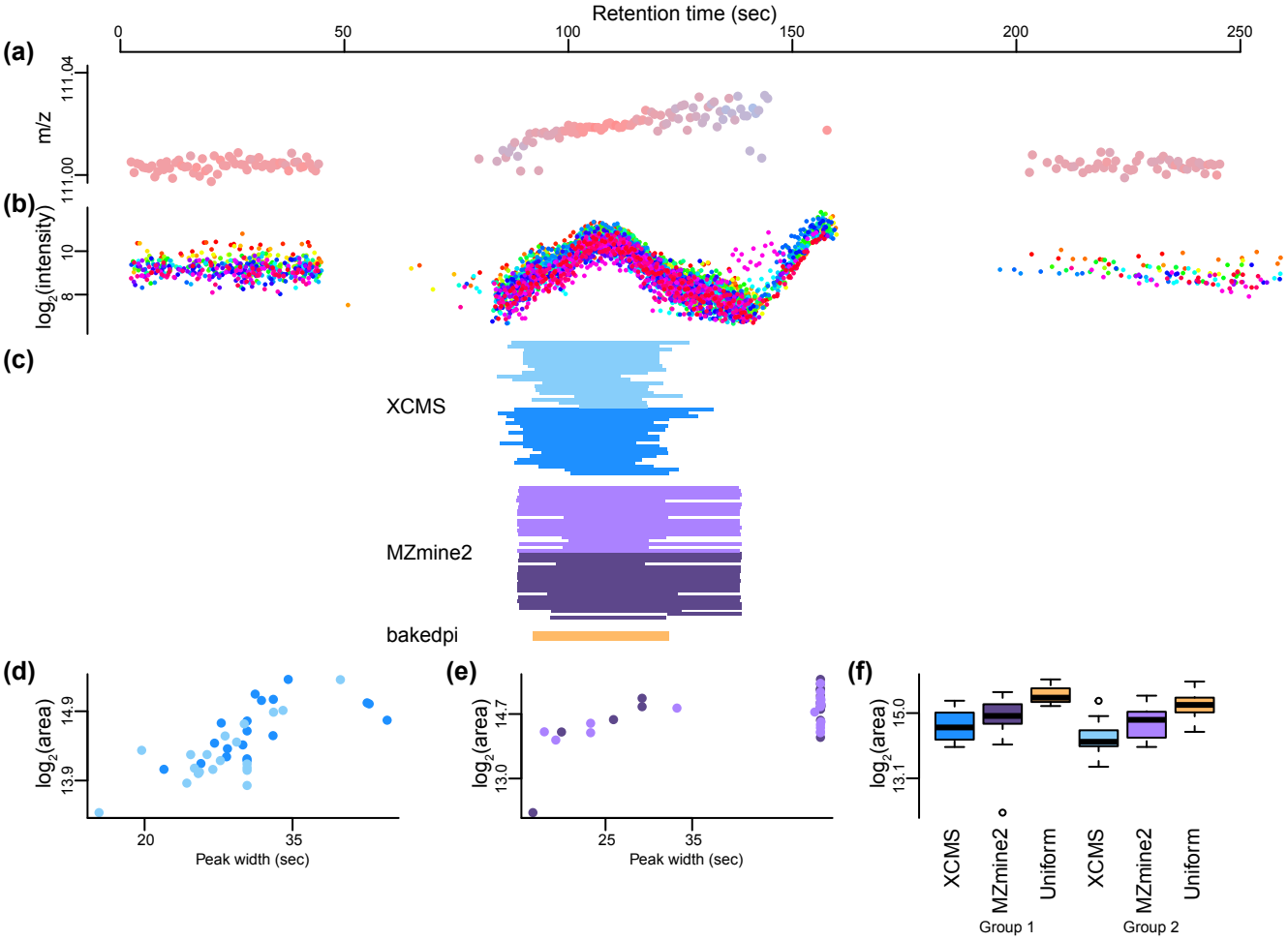
Join aligner

- m/z tolerance: We used 0.005 m/z for the absolute tolerance and the optimized XCMS ppm for the ppm tolerance.
- RT tolerance: We used the maximum peak width from XCMS
- Weight for M/Z and RT: We set these both to 1

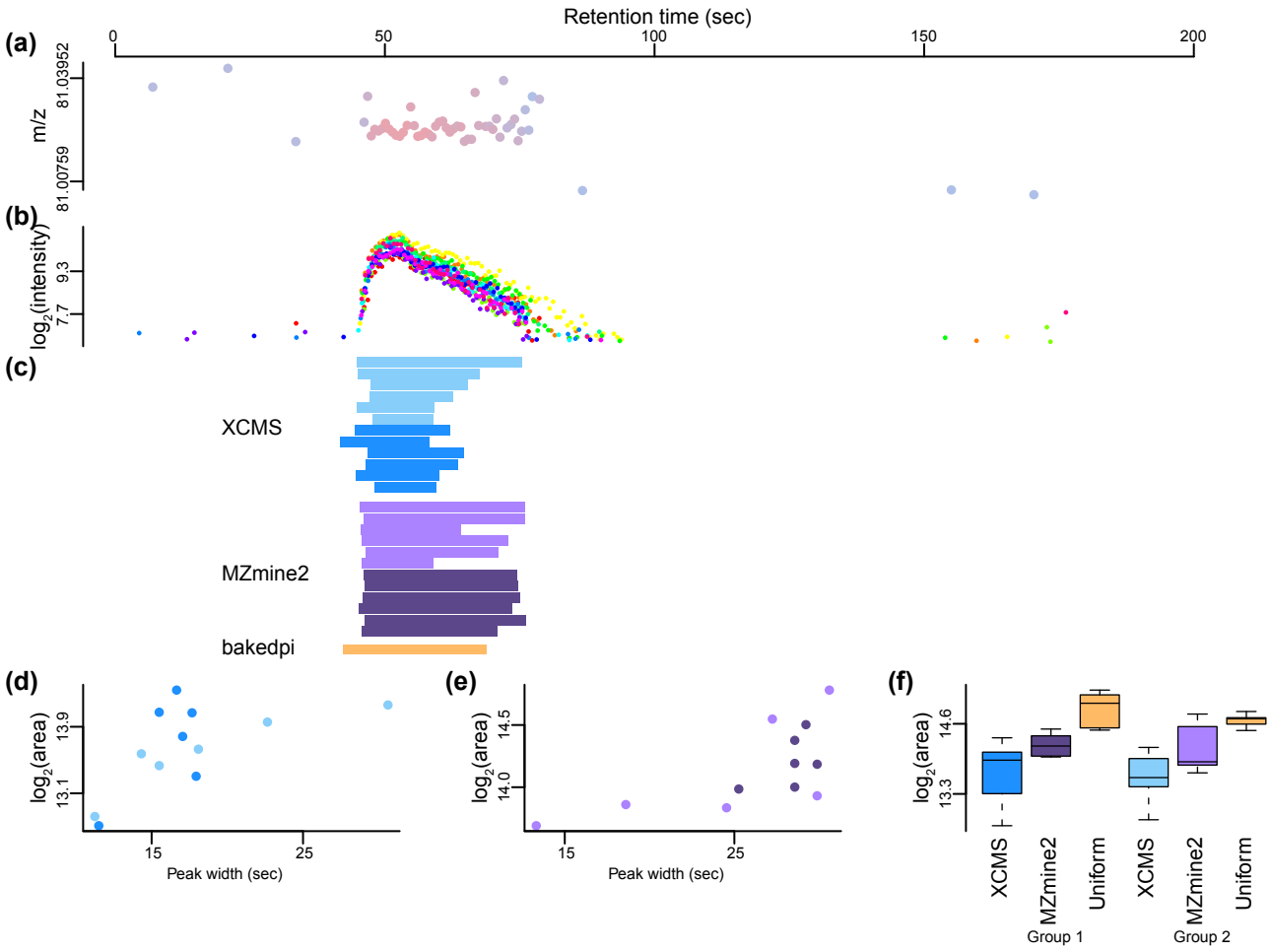
Same-range gap filler

- m/z tolerance: We used 0.005 m/z for the absolute tolerance and the optimized XCMS ppm for the ppm tolerance.

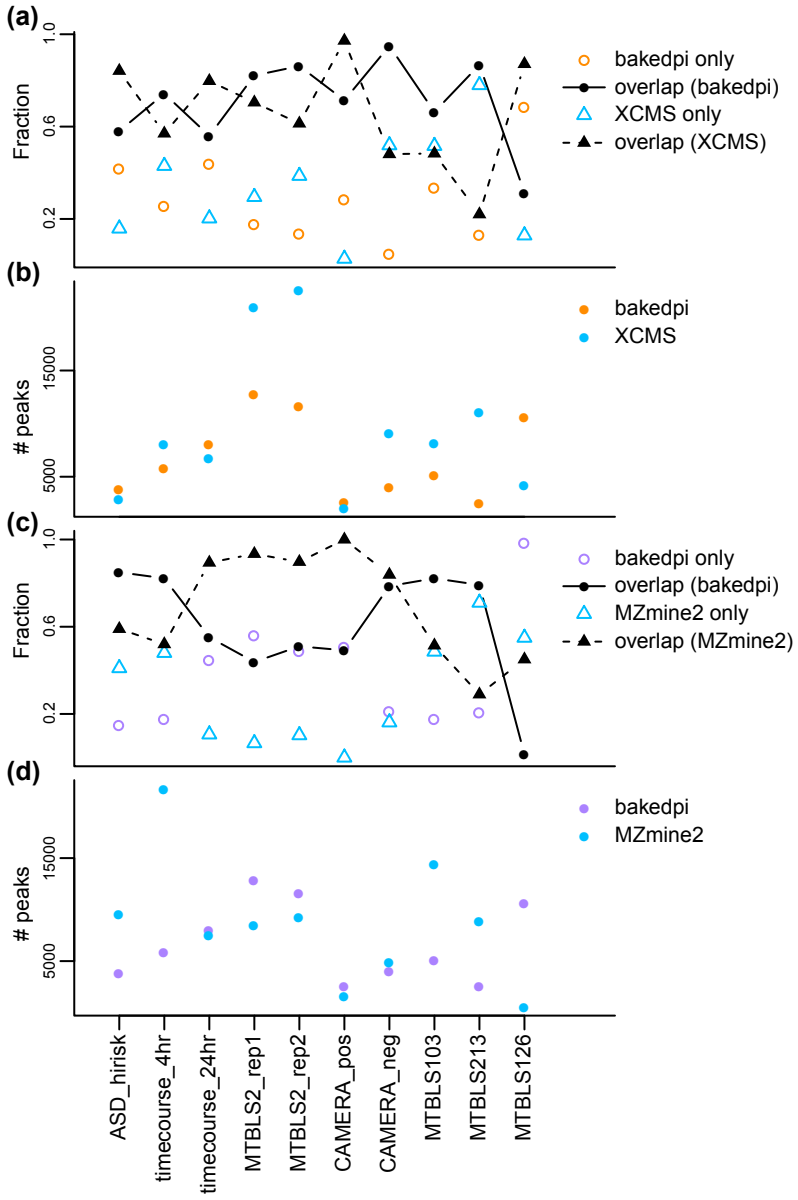
Supplementary Figures



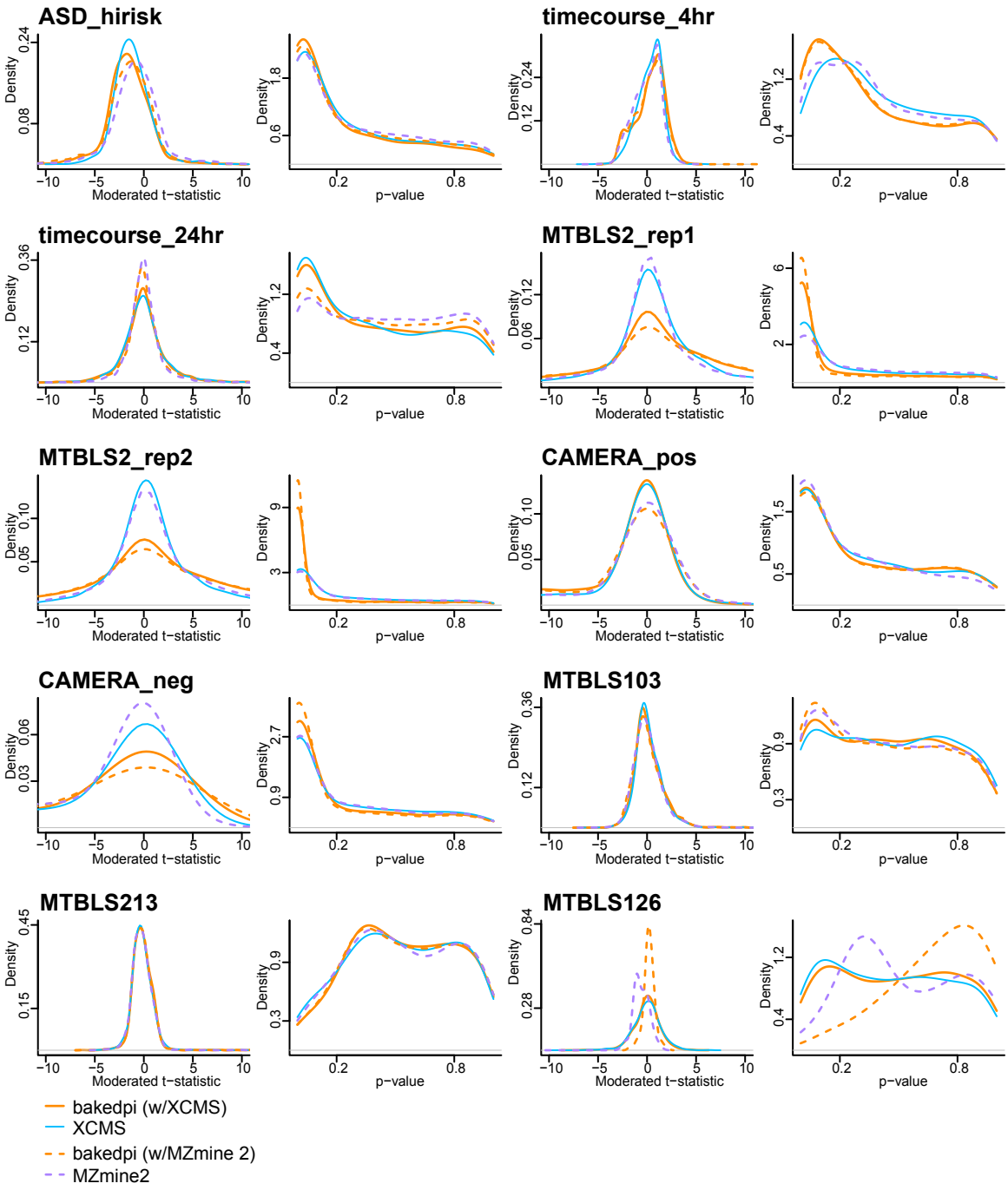
Supplementary Figure S1. Problems with XCMS and MZmine2 processing. Like Figure 1, but from the ASD_hirisk dataset. **(a)** The m/z-RT space surrounding this peak in a single sample, color is used to indicate intensity (red is high). **(b)** Overlaid extracted ion chromatograms from all 40 samples in the experiment. Different colors denote different samples. **(c)** The peak bounds for all samples for XCMS (blue), MZmine2 (purple) and bakedpi (orange; all samples have same bounds). This experiment compares two groups of samples indicated with different color shades. **(d)** XCMS peak quantification vs. peak width. **(e)** Like (d) but for MZmine. **(f)** Distribution of peak quantifications, based on the peak bounds in (c). Substantial heterogeneity in the sample-specific bounds leads to excess variability in the quantifications; this is addressed by using the same RT bound for all samples.



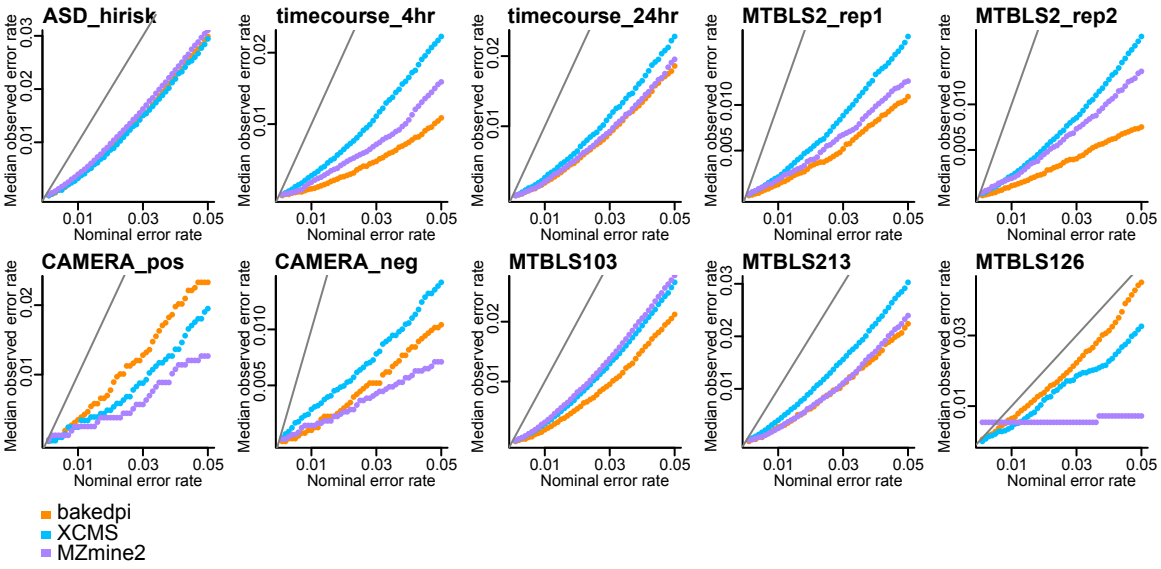
Supplementary Figure S2. Problems with XCMS and MZmine2 processing. As Supplemental Figure S1, depicting an example from the timecourse_4hr dataset.



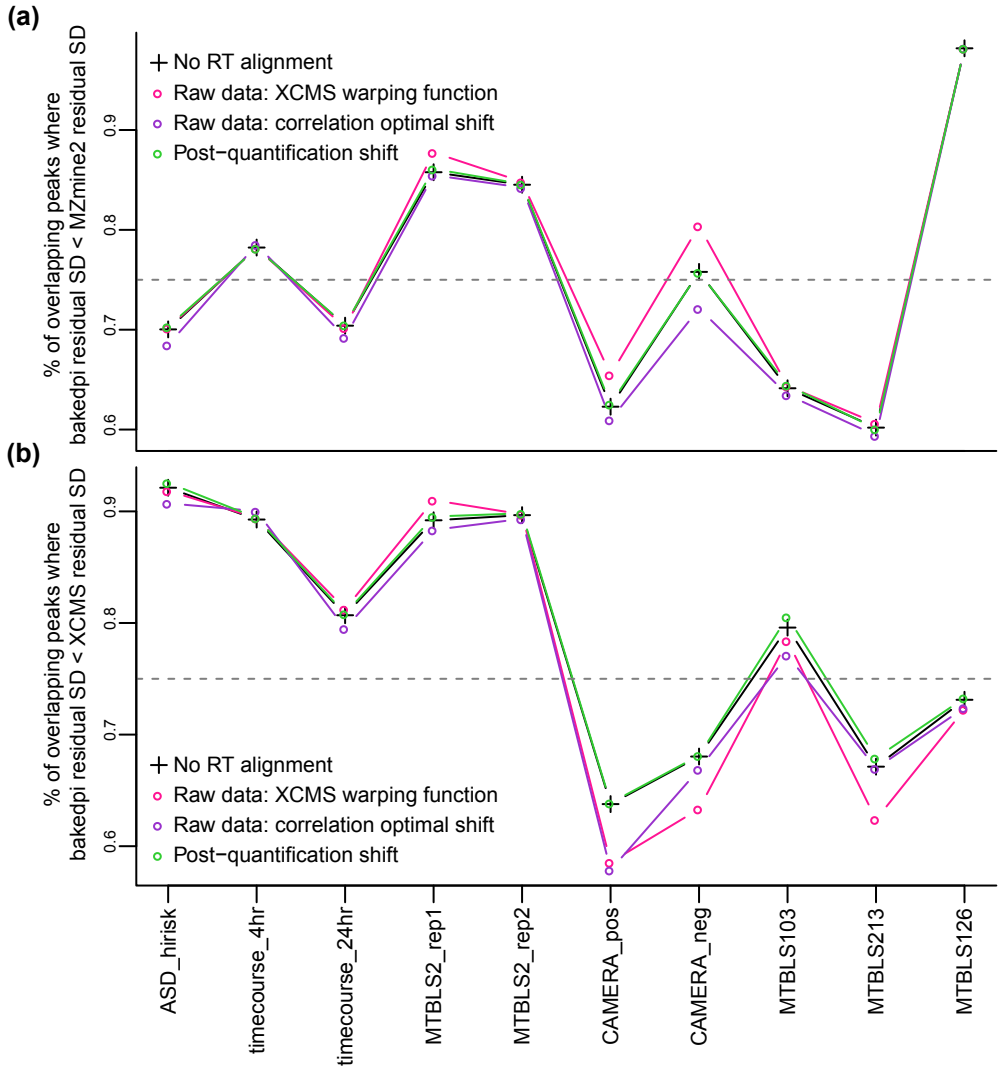
Supplementary Figure S3. Number of peaks called and overlap between methods. (a) The peaks detected by bakedpi are split into two groups: those that are only detected by bakedpi and those that are also detected by XCMS (orange and black circles). XCMS peaks are split similarly (blue and black triangles). **(b)** The number of peaks detected by bakedpi and XCMS. **(c), (d)** Like (a), (b) but for the bakedpi-MZmine2 comparison. In most datasets, bakedpi and the comparison method detect a similar number of peaks, a large percentage of which are found by both methods. Still for nearly all datasets, there is a sizable number of peaks which are only detected by one method.



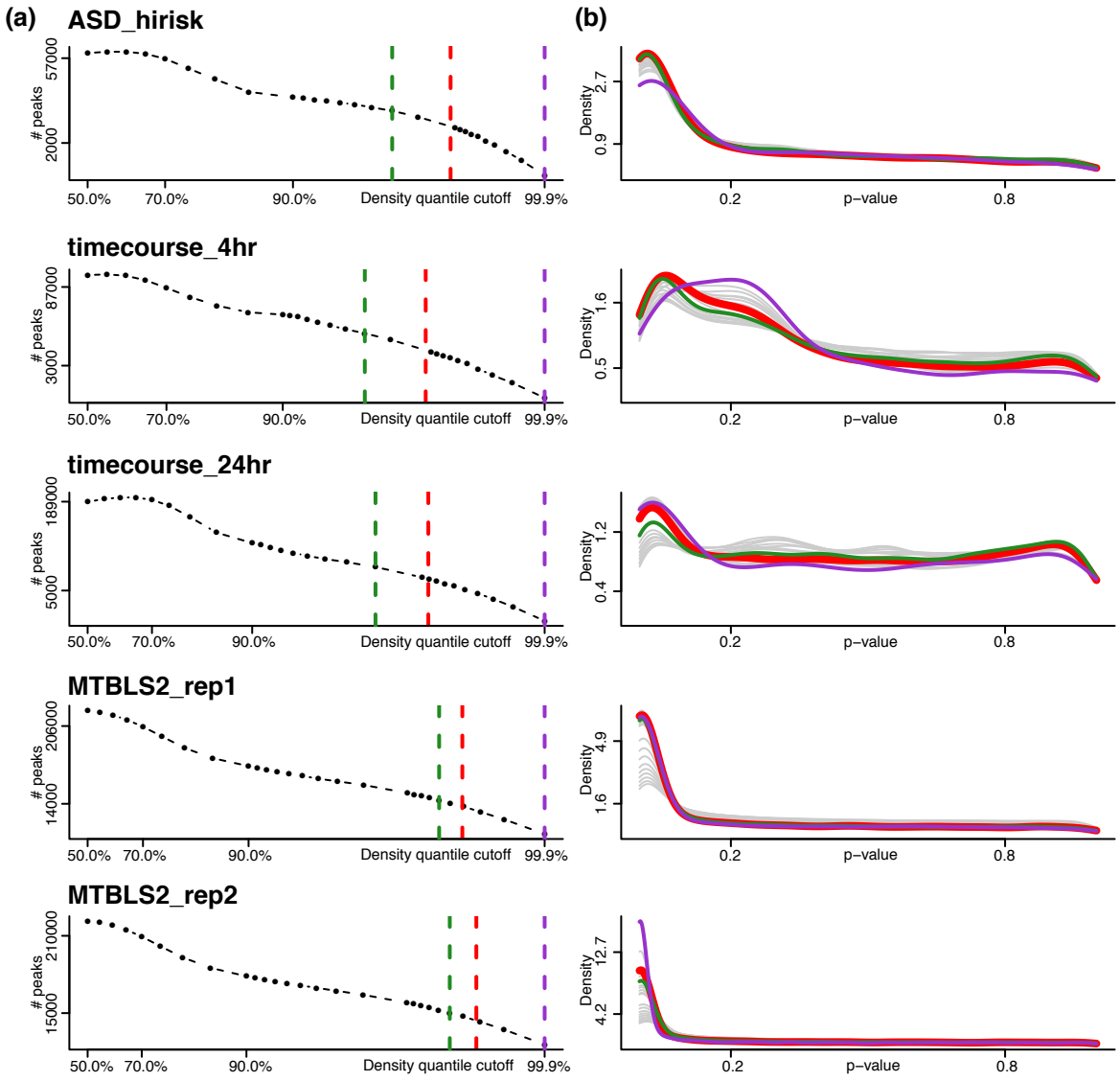
Supplementary Figure S4. Comparison of differential analysis quality in peaks detected by both bakedpi and either XCMS or MZmine2. The `limma` package was used to perform differential abundance analysis on quantifications from bakedpi and XCMS. Shown here are the distributions of the moderated t-statistics and associated p-values for the peaks detected by both bakedpi and XCMS (solid lines) and for the peaks detected by both bakedpi and MZmine2 (dotted lines).



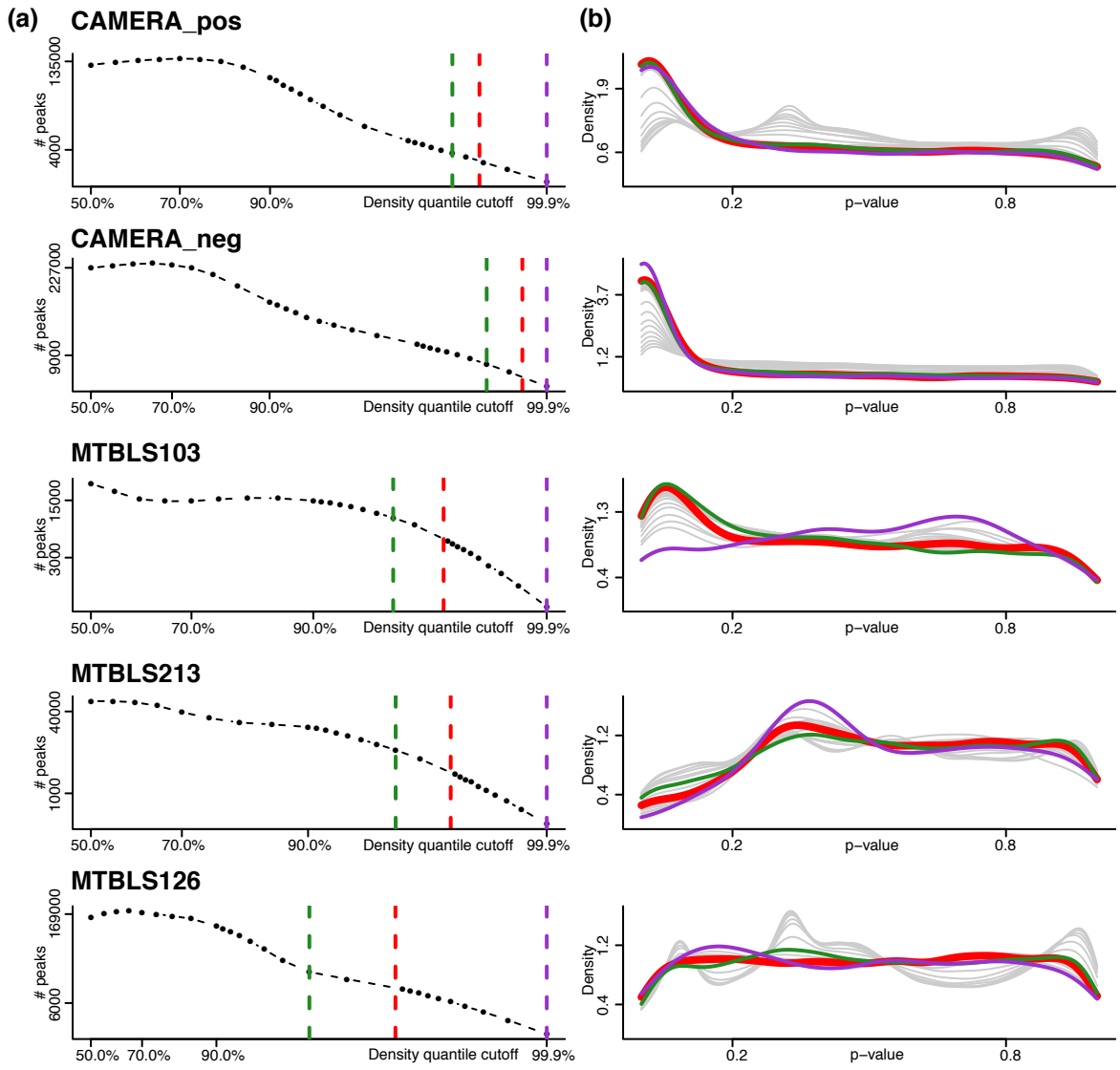
Supplementary Figure S5. bakedpi has more conservative type I error control than XCMS and MZmine2. For each dataset, sample labels were permuted to create null comparisons in which the new permuted groups both had an equal mix of original case and control samples. The median error rate over these null permutations is shown as a function of the nominal error rate. For all datasets, both bakedpi and XCMS are conservative, and for most datasets, bakedpi is as or more conservative than XCMS and MZmine2.



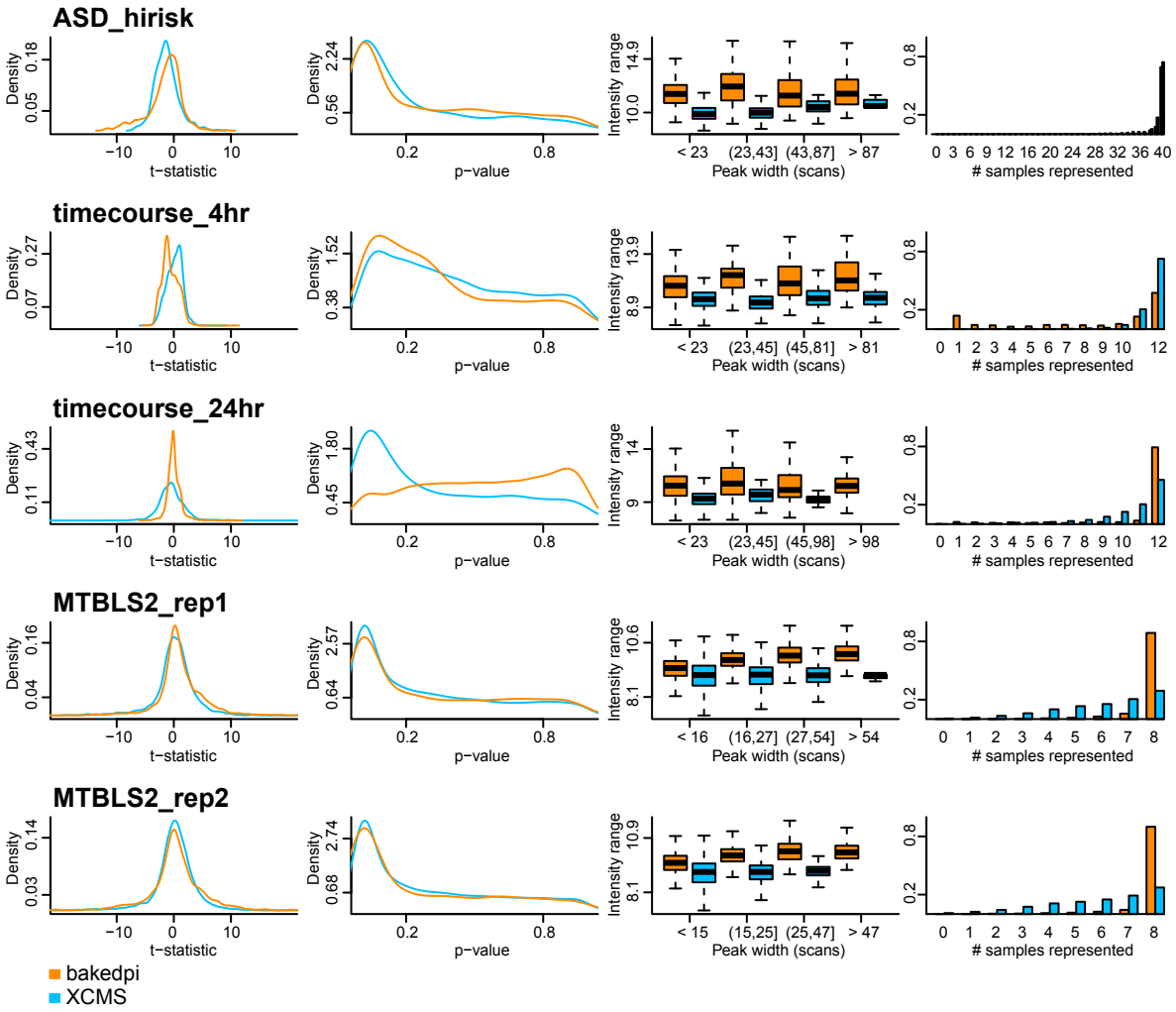
Supplementary Figure S6. Impact of RT alignment. (a) Percentage of peaks overlapping between bakedpi and MZmine2 for which quantification variability is higher in MZmine2 for various RT alignment strategies. **(b)** Like (a) but for XCMS.



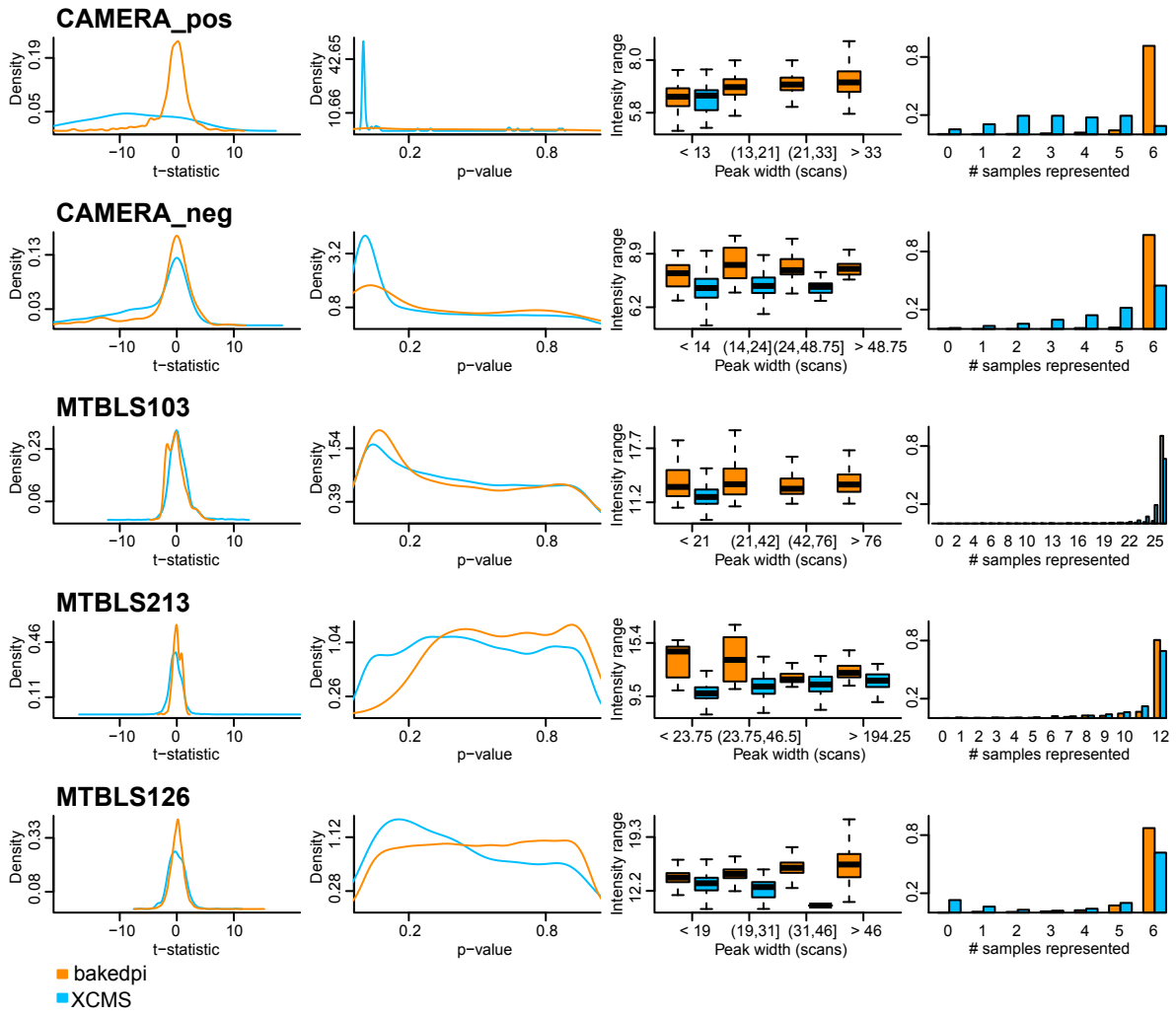
Supplementary Figure S7. Sensitivity of results to density cutoff. (a) Number of peaks detected by bakedpi as a function of the density cutoff. **(b)** The p-value distributions corresponding to the range of cutoffs. Shown in red is the cutoff actually picked by bakedpi. Shown in green and purple are slightly lower and slightly higher cutoffs.



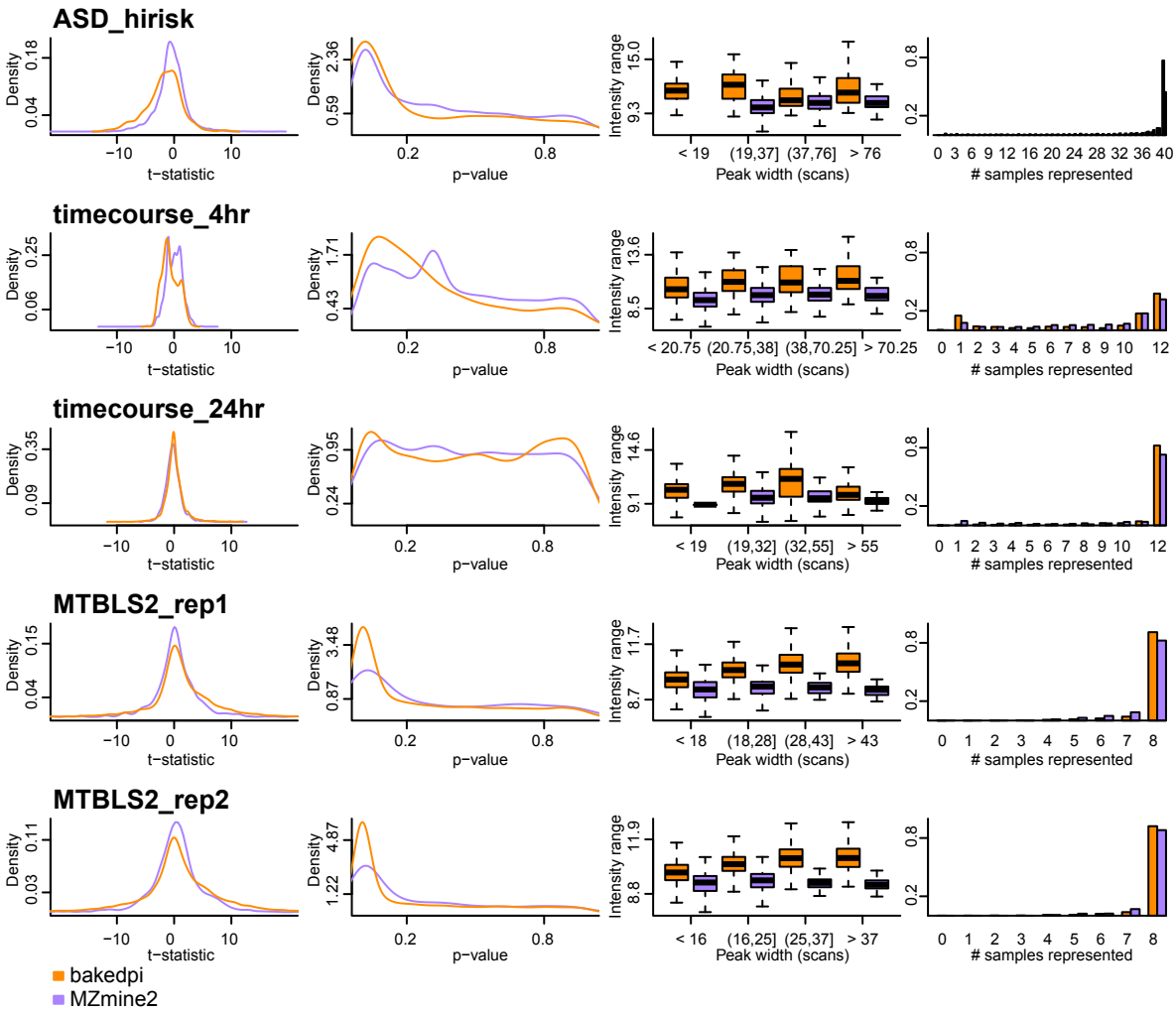
Supplementary Figure S8. Sensitivity of results to density cutoff. As Supplemental Figure S7, but for 5 additional datasets.



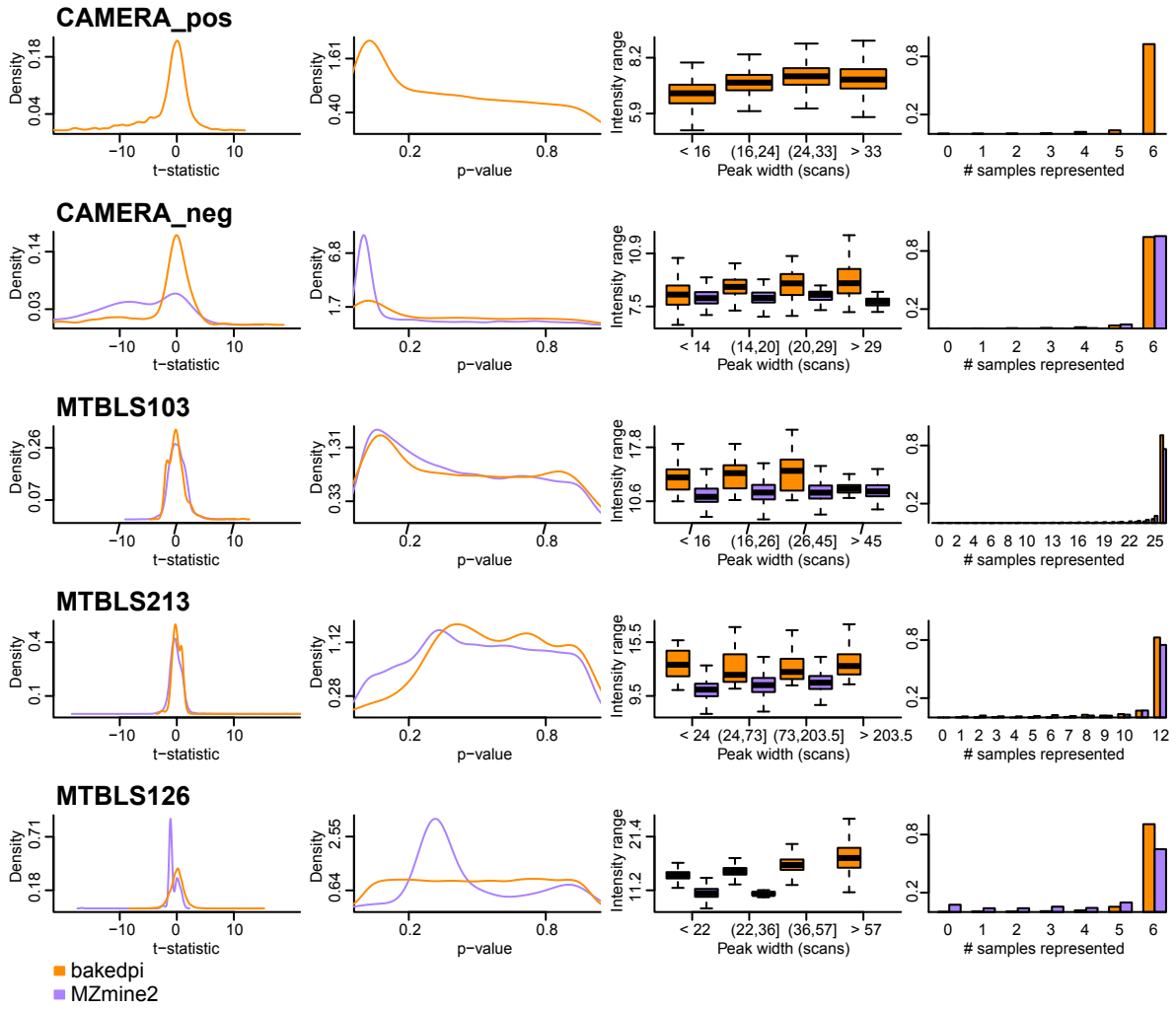
Supplementary Figure S9. Characteristics of peaks that are detected only by one method: bakedpi-XCMS comparison. Columns 1-4 show, respectively, the distribution of t-statistics, p-values, intensity ranges (log2), and number of samples represented for peaks detected only by bakedpi (orange) and only detected by XCMS (blue). The intensity range within a peak is a measure of peak height and is shown as a function of peak width.



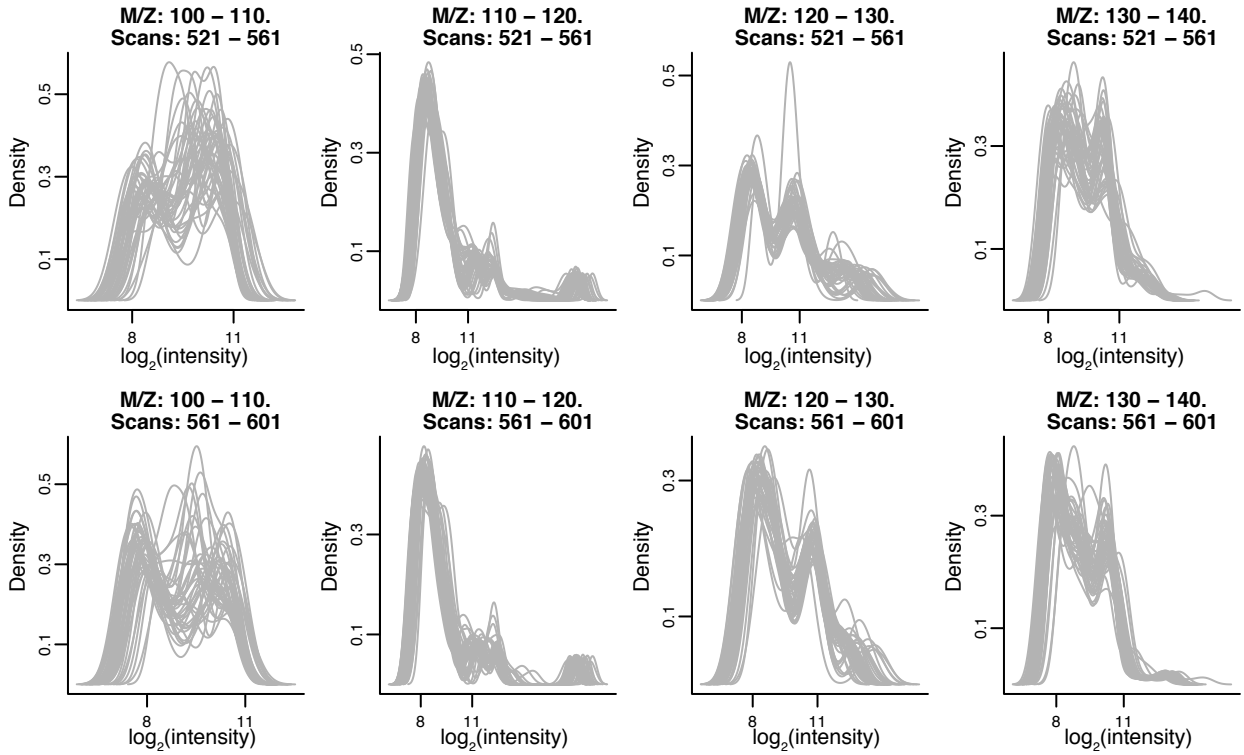
Supplementary Figure S10. Characteristics of peaks that are detected only by one method: bakedpi-XCMS comparison. As Supplementary Figure S9, but for 5 additional datasets.



Supplementary Figure S11. Characteristics of peaks that are detected only by one method: bakedpi-MZmine2 comparison. As with the bakedpi-XCMS comparisons (Supplementary Figures S9 and S10), with the first 5 datasets.



Supplementary Figure S12. Characteristics of peaks that are detected only by one method: bakedpi-MZmine2 comparison. As Supplementary Figure S11, but for 5 additional datasets.



Supplementary Figure S13. Region-specific intensity distributions. Each plot depicts the intensity distribution over a single grid region in the m/z-RT space, for the ASD_hirisk dataset. Each line corresponds to a single sample.