

# Supporting information

## Deep Learning assisted Peak Curation for large scale LC-MS Metabolomics

Yoann Gloaguen<sup>1,2,3</sup>, Jennifer A. Kirwan<sup>1,3</sup>, Dieter Beule<sup>2,3\*</sup>

<sup>1</sup>Berlin Institute of Health @ Charité, Metabolomics Platform, 10178 Berlin, Germany

<sup>2</sup>Berlin Institute of Health @ Charité, Core Unit Bioinformatics, 10178 Berlin, Germany

<sup>3</sup>Max Delbrück Center for Molecular Medicine in the Helmholtz Association, 13125 Berlin, Germany

\*Corresponding author: [dieter.beule@bih-charite.de](mailto:dieter.beule@bih-charite.de)

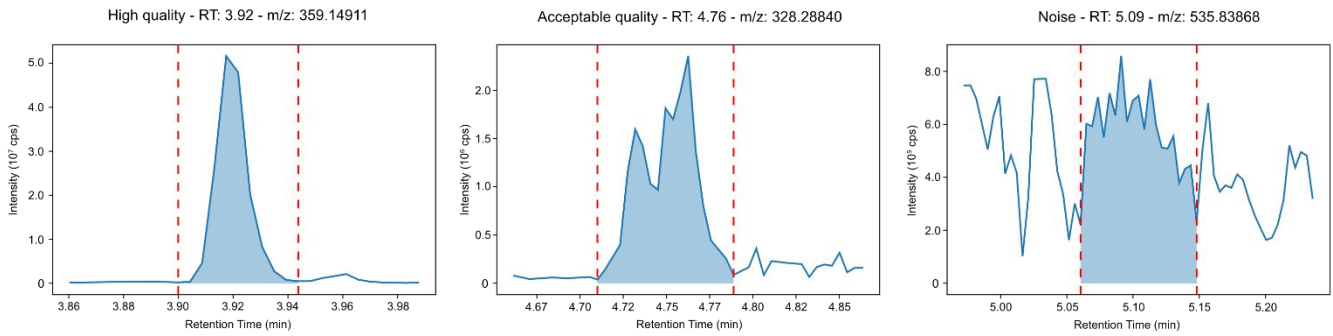


Figure S1: Example of CS peaks detected by MZmine on dataset 2 and predicted as High quality, Acceptable quality and Noise by NeatMS.

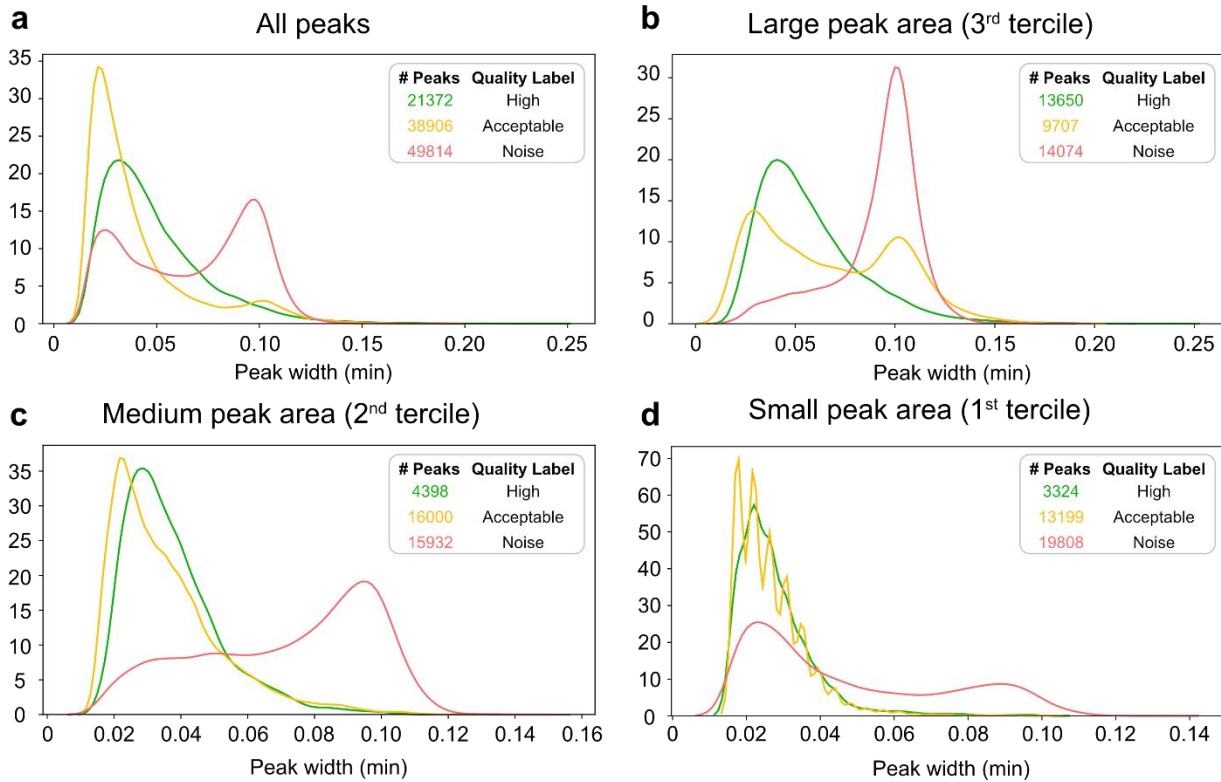


Figure S2: Density plots representing the peak width distribution of the 3 peak classes of dataset 1 analysed using MZmine and NeatMS TL model. a. Width distribution of all peaks. b. Width distribution of peaks from the upper tertile of peak area. c. Width distribution of peaks from the intermediate peak area tertile. d. Width distribution of peaks from the lower peak area tertile.

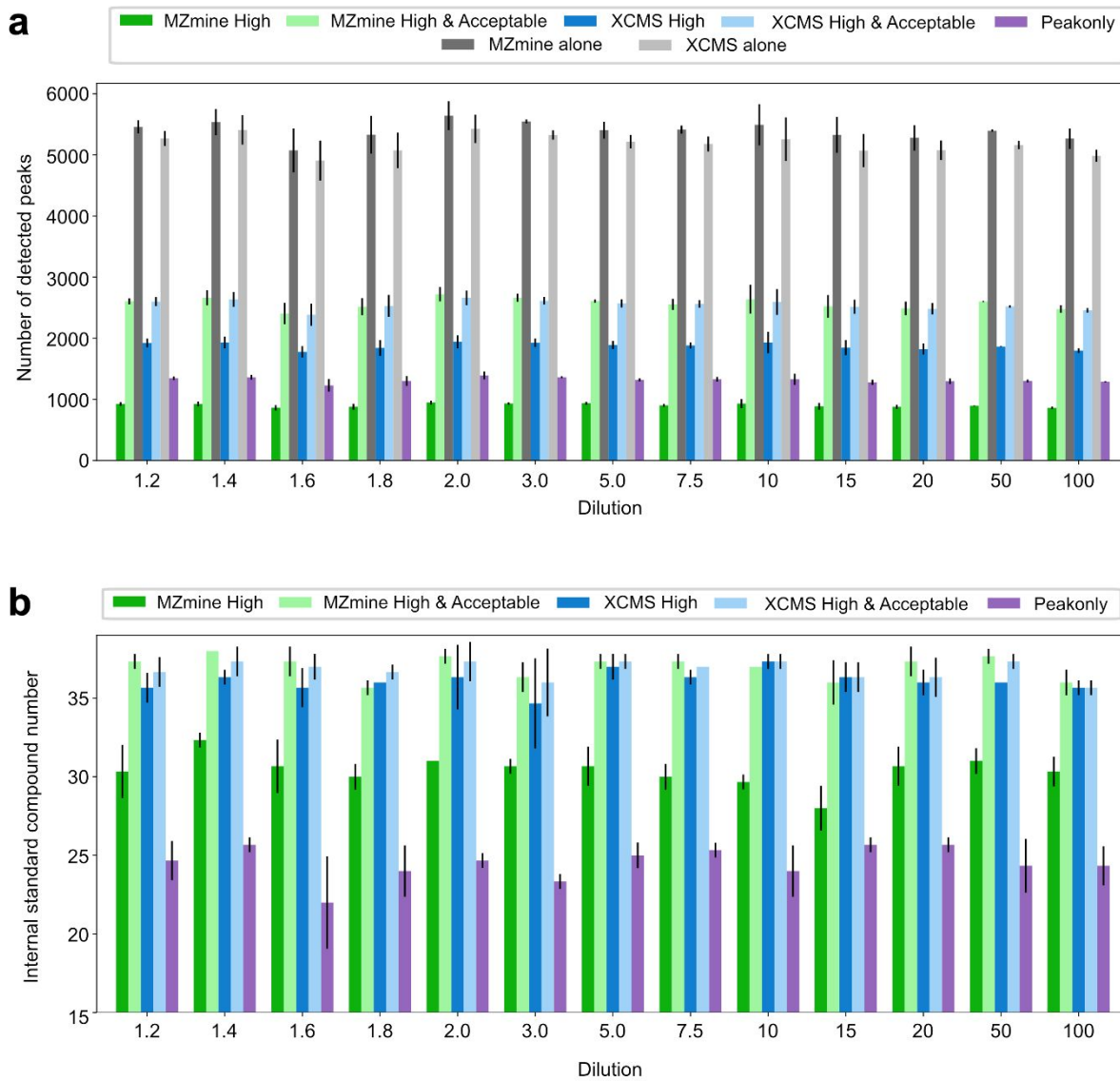


Figure S3: Peak number and internal standard recovery for different tools on the different dilution points of dataset 2. a. Average number of peaks for each dilution point before and after using NeatMS. Only the high quality and acceptable quality classes of NeatMS are displayed here. The difference between the total number of peaks detected by MZmine and XCMS and the number of peaks included in NeatMS high and acceptable quality corresponds to peaks predicted as Noise and peaks rejected by the minimum scan number filter (set to the default value of 5). b. Number of (non diluted) internal standard compounds recovered by the different tools and the details of their predicted classes for each dilution point.

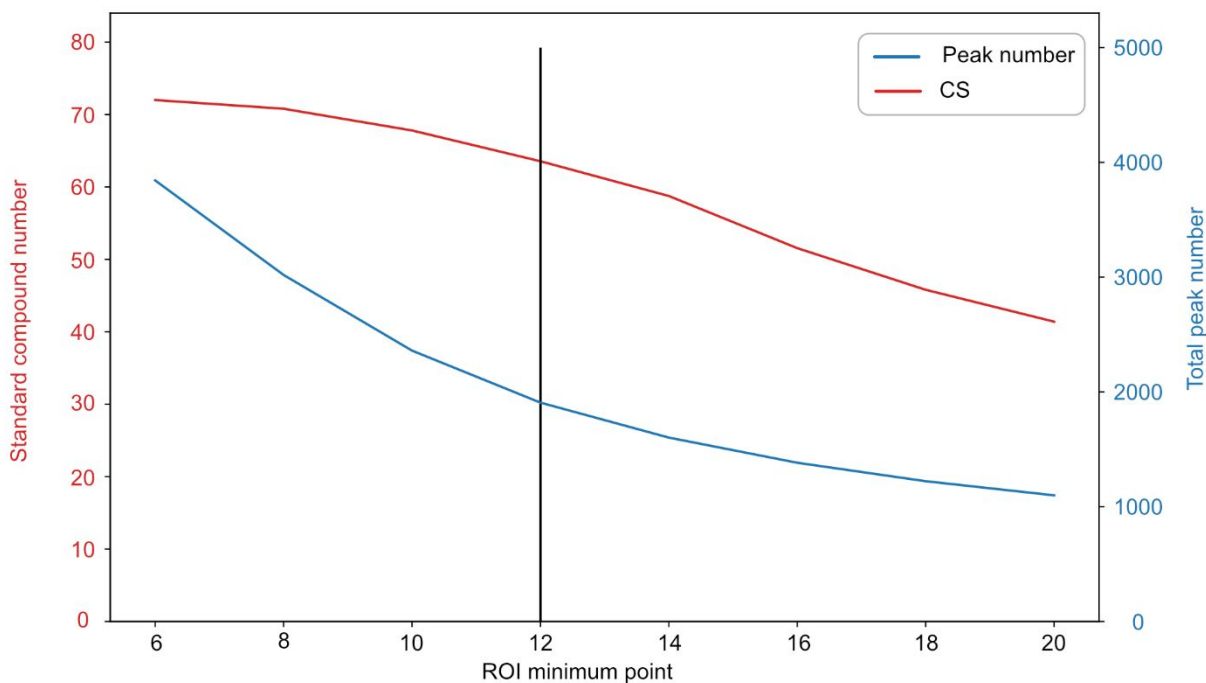


Figure S4: Average detected peak number and standard compound recovered by peakonly on dataset 1 when modulating peakonly “ROI minimum points” parameter. As previously described by the authors of peakonly, we found that lowering the ROI minimum points parameter significantly increases the number of reported noise peaks. A similar effect can be observed with peakonly “Peak minimum points” parameter (not shown). We conclude that the recommendations of peakonly authors are a good compromise between CS sensitivity and false positive peak detection. The vertical line (in black) represents the parameter that we have used for comparison.

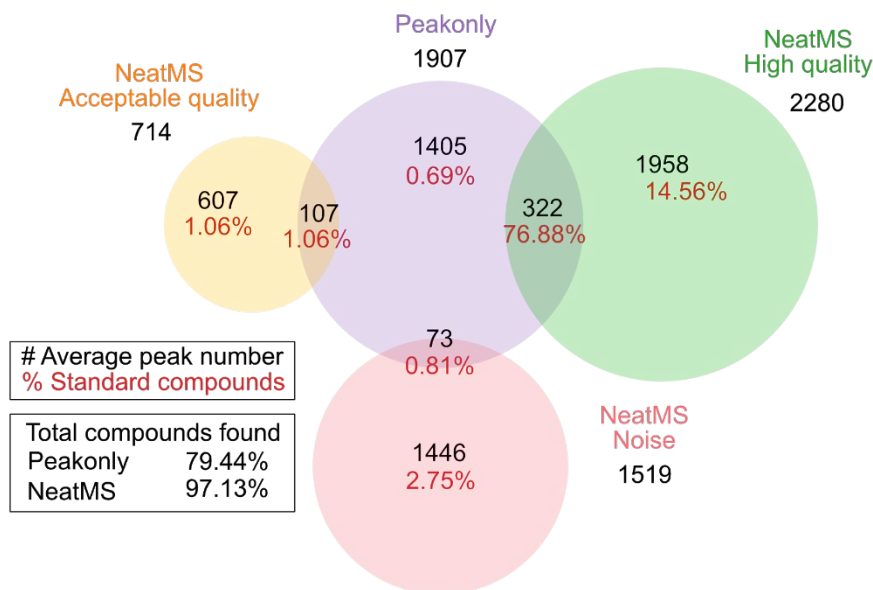


Figure S5: Venn diagram comparing peakonly and the combination XCMS with NeatMS TL model. Numbers are averages over the 20 samples of dataset 1: total number of detected peaks (black), percent of recovered CS (red).

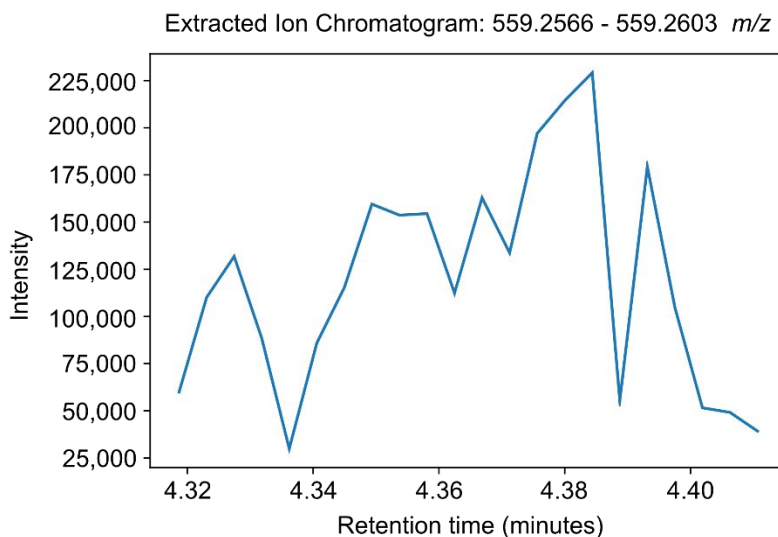


Figure S6: Example of an extracted ion chromatogram of a peak reported by MZmine and classified as Noise by NeatMS matching a CS in dataset 1.

**Table S1: Parameters of the different tools used for dataset 1 and 2. Only non default parameters are shown. Peakonly stable and published version 0.1 was used through command lines. XCMS version 3.10.0 was used in R 4.0.1. MZmine was used under version 2.53. NeatMS was used under version 0.6 with default parameters only.**

Parameter	Value
MZmine (Dataset 1 & 2)	
ADAP chromatogram builder	
Min group size in # of scans	5
Group intensity threshold	$5 \times 10^2$
Min highest intensity	$1 \times 10^3$
m/z tolerance (m/z)	$1 \times 10^{-2}$
Wavelets (ADAP) chromatogram deconvolution	
S/N threshold	10
S/N estimator	Intensity window SN
Min feature height	$1 \times 10^3$

Peak duration range	0.02 - 1.0
RT wavelet range	0.001 - 0.05
XCMS CentWave (Dataset 1)	
Min-Max peak width	3-85.16
ppm	17
mzdiff	-0.01145
XCMS CentWave (Dataset 2)	
Min-Max peak width	3-81
ppm	11.75
mzdiff	-0.016
Peakonly (Dataset 1 & 2)	
Delta mz	0.01
ROI minimum points	12
Peak minimum points	6

**Table S2: Performance of peakonly with NeatMS using dataset 1 and the two models: Average number of peaks across 20 samples, average percentages of the 80 SC recovered using peakonly and different models. The input row shows the results returned by peakonly alone, other rows show the details of the three peak classes given by NeatMS. The total number of peaks after classification is smaller than the input due to the application of a minimum scan number filter that NeatMS uses (default value of 5 is used).**

		Peakonly data with NeatMS TL model	Peakonly data with NeatMS PT model
Peak number	Input	1907	1907
	Classified	1724	1724
	High Quality	457	730

	Acceptable Quality	747	684
	Noise	520	310
CS found	Input & classified	79.44%	79.44%
	High Quality	44.75%	74.69%
	Acceptable Quality	33.06%	3.81%
	Noise	1.44%	0.75%

**Table S3: NeatMS results using TL model on dataset 1 analysed using MZmine with and without the gapfilling step. Peak number: average number of peaks across 20 samples. CS found: average percentages of the 80 CS recovered.**

		MZmine without gapfilling	MZmine with gapfilling
Peak number	Input	6977	21916
	Classified	5505	14985
	High quality	1069	1788
	Acceptable	1945	2487
	Noise	2491	10710
CS found	Input & classified	94.25%	95.00%
	High quality	79.31%	81.81%
	Acceptable	11.25%	9.63%
	Noise	3.69%	3.56%

**Table S4: Confusion matrix of DNN and NeatMS number of peaks per class on dataset 1 analyzed using MZmine with gapfilling. TL model was used for NeatMS. Numbers are averaged over 20 samples. No match represents peaks without an exclusive matching peak or without any matching peak at all. Column entries represent NeatMS results, row entries represent DNN results.**

	High quality	Acceptable	Noise	No Match	Total
Good	832	1351	1188	5292	8663

Bad	305	346	5200	10770	16621
No Match	651	790	4322	/	
Total	4275		10710		

**Table S5: Confusion matrix of DNN and NeatMS CS recovery per class on dataset 1 analyzed using MZmine with gapfilling. TL model was used for NeatMS. Numbers are averaged over 20 samples. No match represents CS missed by the specific method but found using the other. Column entries represent NeatMS results, row entries represent DNN results.**

	High quality	Acceptable	Noise	No Match	Total
Good	73.31%	7.56%	2.69%	1.44%	85.00%
Bad	1.31%	0.25%	0.81%	0.12%	2.49%
No Match	7.20%	1.81%	0.06%	/	
Total	91.44%		3.56%		

**Table S6: Confusion matrix of MetaClean and NeatMS number of peaks per class on dataset 1 analyzed using XCMS with alignment and gapfilling. TL model was used for NeatMS. Numbers are averaged over 20 samples. No match represents peaks without an exclusive matching peak or without any matching peak at all. Column entries represent NeatMS results, row entries represent MetaClean results.**

	High quality	Acceptable	Noise	No Match	Total
Pass	1194	356	508	3464	5522
Fail	1079	410	1140	4940	7569
No Match	1024	282	2204	/	
Total	4345		3852		

**Table S7: Confusion matrix of MetaClean and NeatMS CS recovery per class on dataset 1 analyzed using XCMS with alignment gapfilling. TL model was used for NeatMS. Numbers are averaged over 20 samples. No match represents CS missed by the specific method but found using the other. Column entries represent NeatMS results, row entries represent MetaClean results.**

	High quality	Acceptable	Noise	No Match	Total
Pass	33.13%	0.94%	1.44%	0.05%	35.56%
Fail	39.25%	1.12%	1.25%	6.43%	43.31%
No Match	18.74%	0.5%	0.06%	/	
Total	93.68%		2.75%		