

# MS-DAP: Mass Spectrometry Downstream Analysis Pipeline

version: beta 0.2.8.1 <https://github.com/ftwkoopmans/msdap/>

## Contents

<b>1 Quality control</b>	<b>2</b>
1.1 DIA confidence score distributions . . . . .	2
1.2 number of peptides and proteins . . . . .	26
1.3 data completeness . . . . .	34
1.4 abundance distributions . . . . .	36
1.5 retention time . . . . .	45
1.6 variation among replicates . . . . .	152
1.7 PCA . . . . .	175
<b>2 Differential abundance analysis</b>	<b>191</b>
2.1 control_Berlin vs AD_Berlin . . . . .	191
2.2 control_Magdeburg vs AD_Magdeburg . . . . .	197
2.3 control_Sweden vs AD_Sweden . . . . .	203
2.4 control_Berlin,control_Magdeburg... vs AD_Berlin,AD_Magdeburg,AD_Sweden . . . . .	209
<b>3 Summary of differential testing</b>	<b>215</b>
<b>4 log</b>	<b>216</b>
<b>5 R command history</b>	<b>218</b>
<b>6 R session info</b>	<b>220</b>

# 1 Quality control

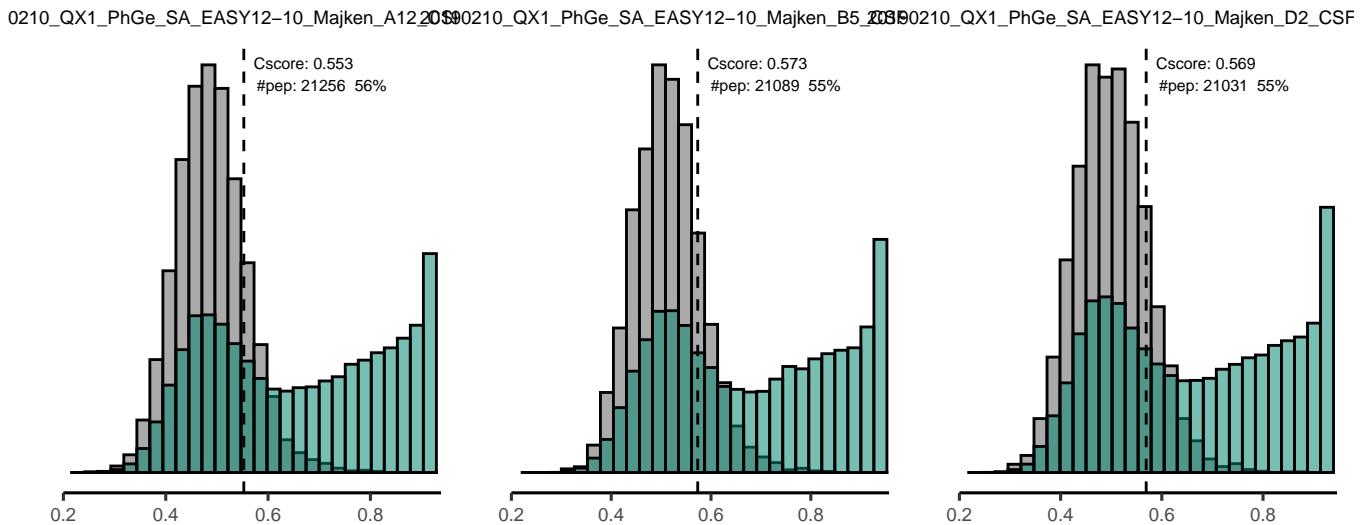
The quality control figures in this section enable you to investigate reproducibility and global clustering of samples by visualizing:

- number of peptides/proteins detected in each sample
- dataset completeness
- local effects in HPLC peptide retention time per sample
- reproducibility of peptide quantification among replicates
- PCA of all samples to visualize clustering

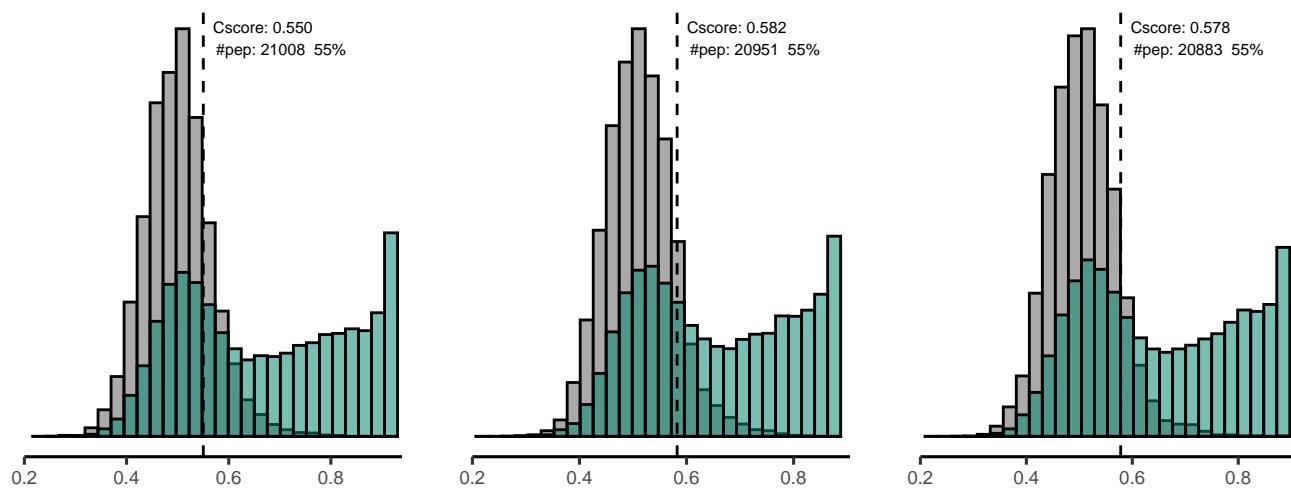
The first set of quality control figures describes individual samples, thereafter group-level quality metrics are described and finally sample clustering is used to highlight structure in the entire dataset.

## 1.1 DIA confidence score distributions

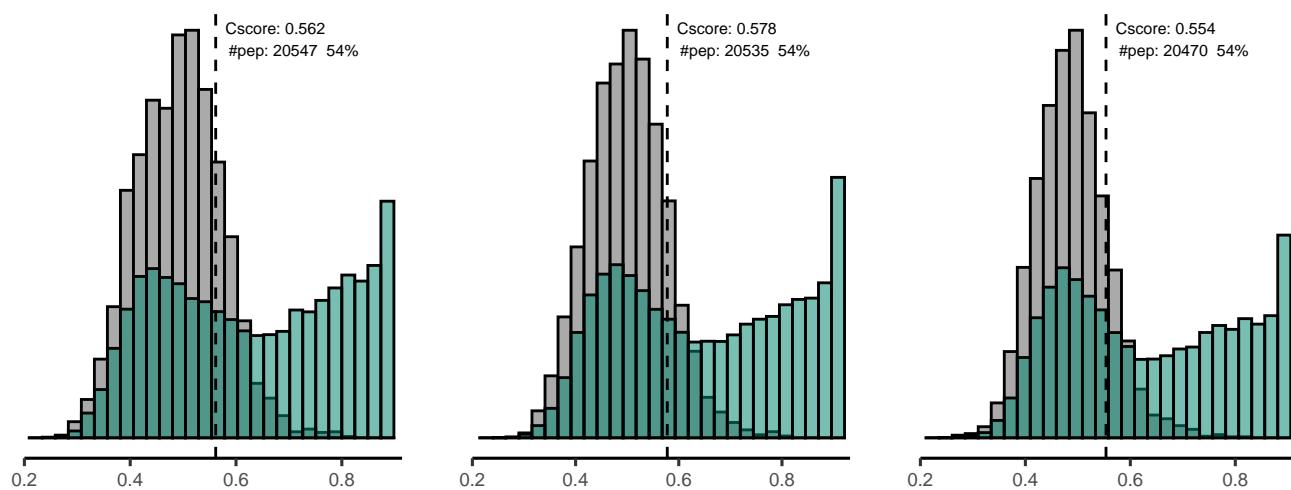
DIA data was used as input for this analysis. Below histograms visualize both target (green) and decoy (grey) cscores in each sample, indicating how confident the input software was in the identification of peptides from the spectral library in the raw DIA data. Samples are ordered by the number of precursors quantified at q-value confidence threshold 0.01. At this threshold, the respective cscore and number of peptides is shown.



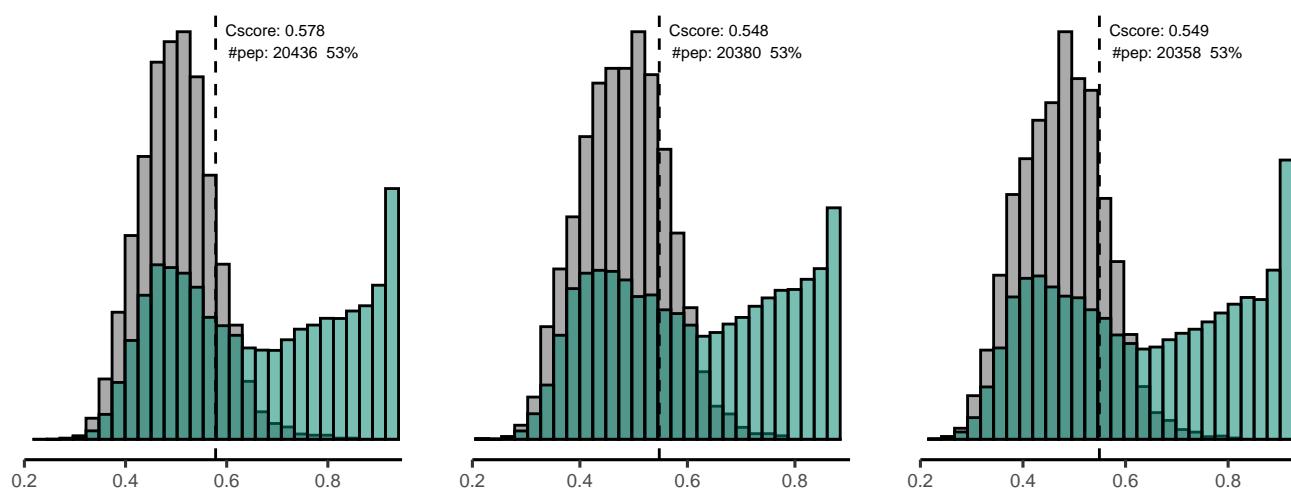
\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_D6\_129\_QX1\_PhGe\_SA\_EASY12-10\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD12910618\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD1xS1fM1.



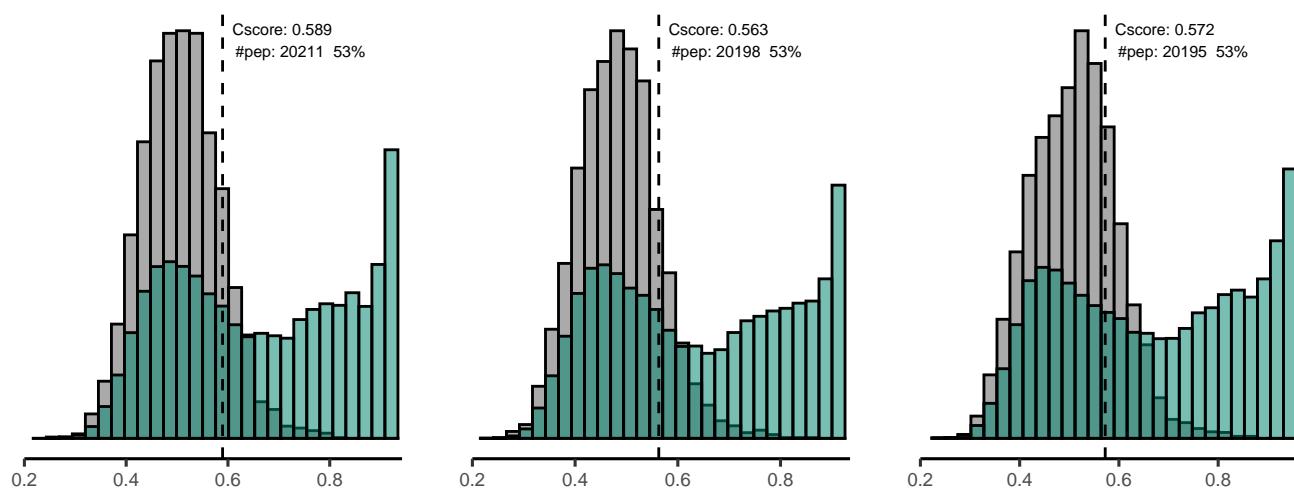
)20210\_QX1\_PhGe\_SA\_EASY12-10\_Majken\_A2\_20590210\_QX1\_PhGe\_SA\_EASY12-10\_Majken20590210\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_D6\_129



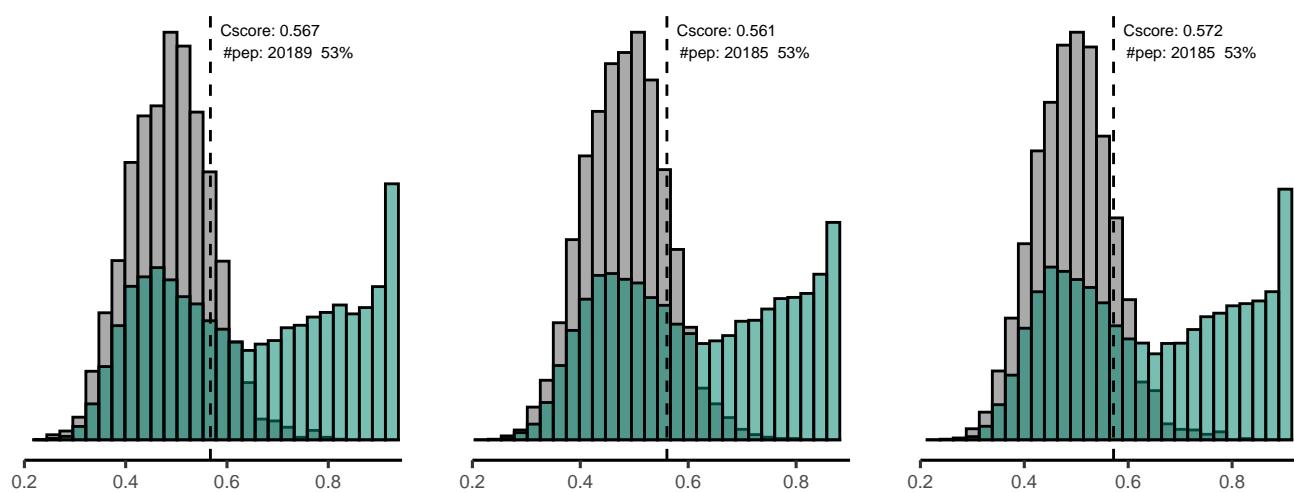
)20210\_QX1\_PhGe\_SA\_EASY12-10\_Majken\_C10\_20590210\_QX1\_PhGe\_SA\_EASY12-10\_Majken\_B7\_20590210\_QX1\_PhGe\_SA\_EASY12-10\_Majken\_C2\_CSF.



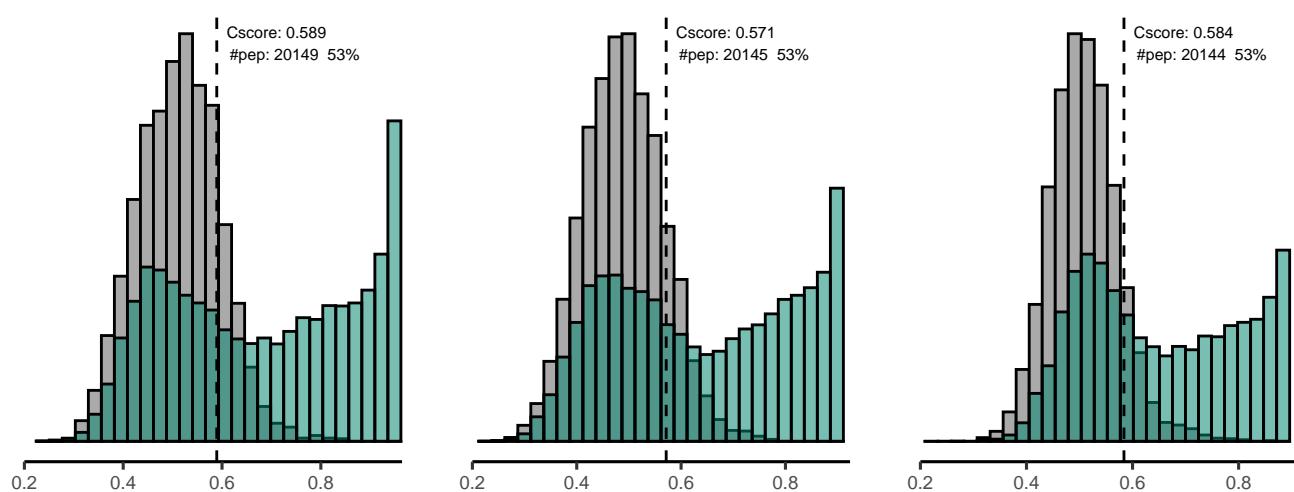
\\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch120190210\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_G120190210\_QX1\_PhGe\_SA\_EASY12-10\_Majken\_D1\_CSF



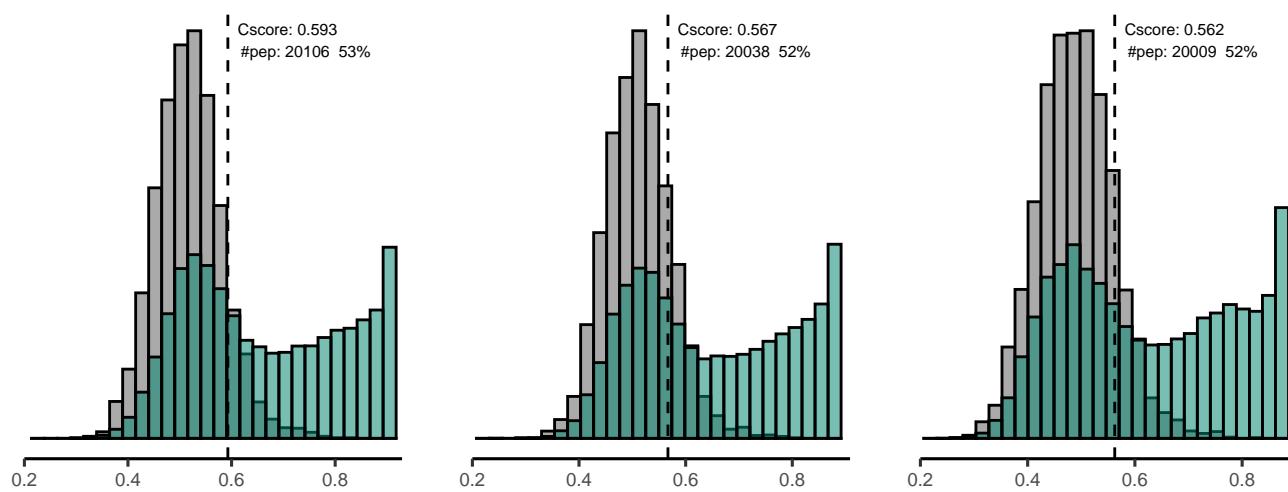
\\_QX1\_PhGe\_SA\_EASY12-10\_Majken\_B8\_20190210\_QX1\_PhGe\_SA\_EASY12-10\_Majken\_20190210\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_D11\_129



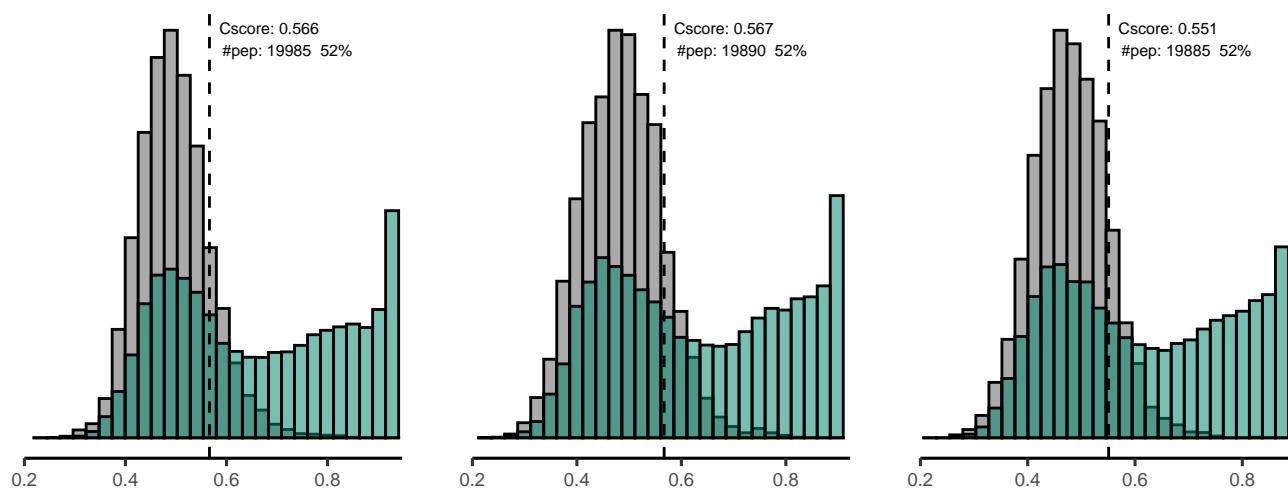
\\_QX1\_PhGe\_SA\_EASY12-10\_Majken\_A3\_20190210\_QX1\_PhGe\_SA\_EASY12-10\_Majken\_A5\_20190218\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD1xS1fm1.



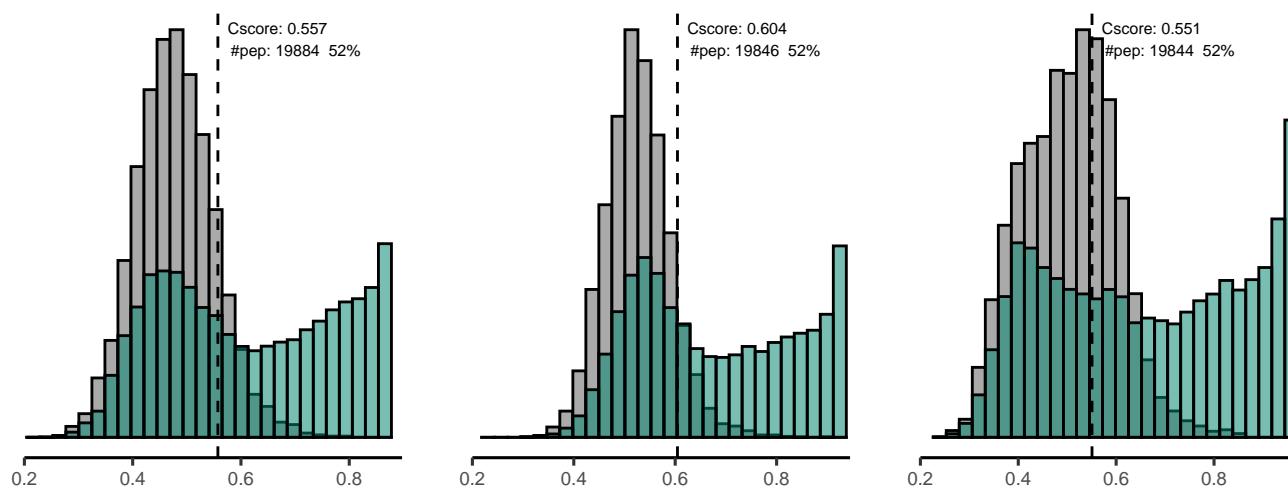
18\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD1xS1028\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_C01028\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_D4\_



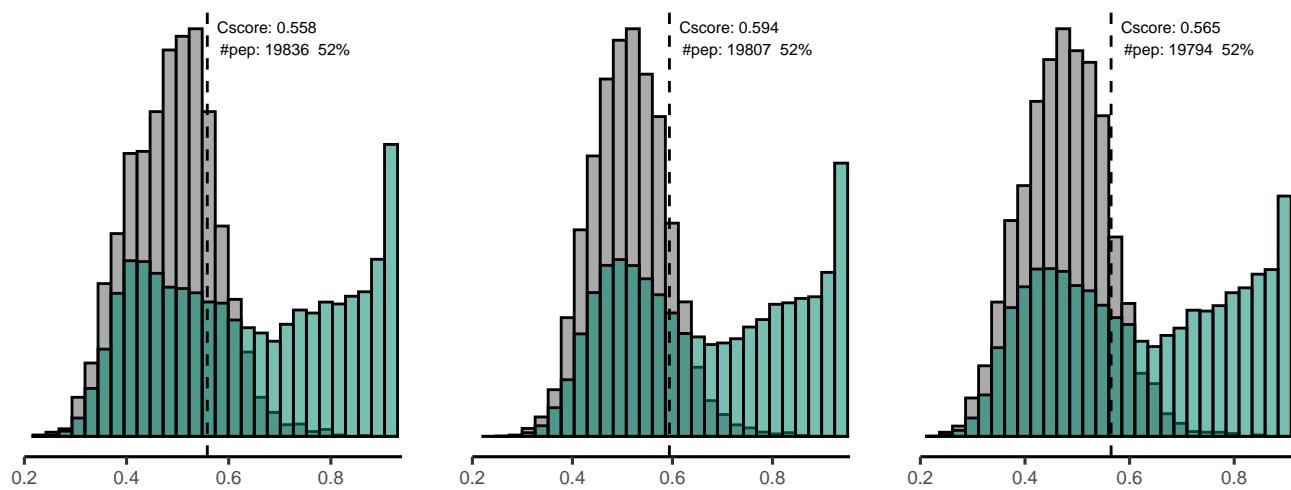
\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_201028\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_A82090210\_QX1\_PhGe\_SA\_EASY12-10\_Majken\_B4\_CSF



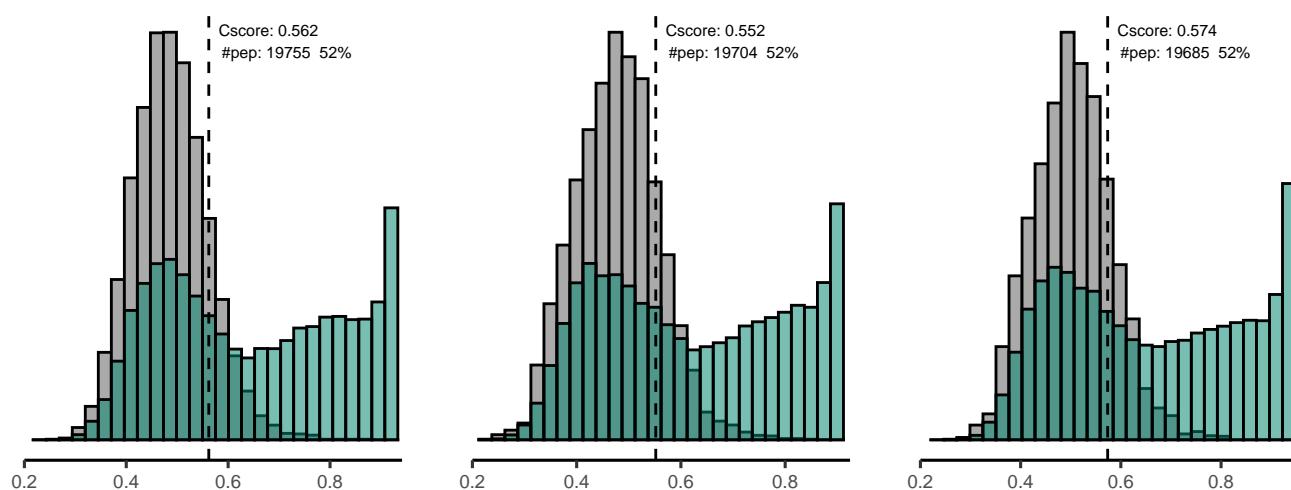
\_0210\_QX1\_PhGe\_SA\_EASY12-10\_Majken\_A9168718\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD1xS1028\_QX1\_PhGe\_SA\_EASY12-10\_Majken\_D6\_CSF



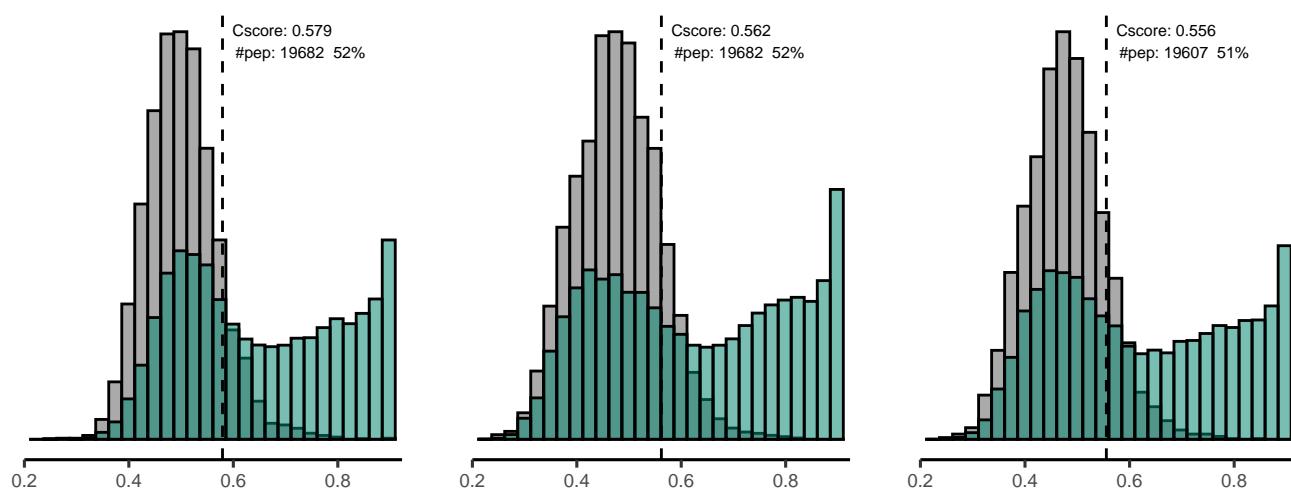
1228\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_20190228\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_20190210\_QX1\_PhGe\_SA\_EASY12-10\_Majken\_D3\_CSF



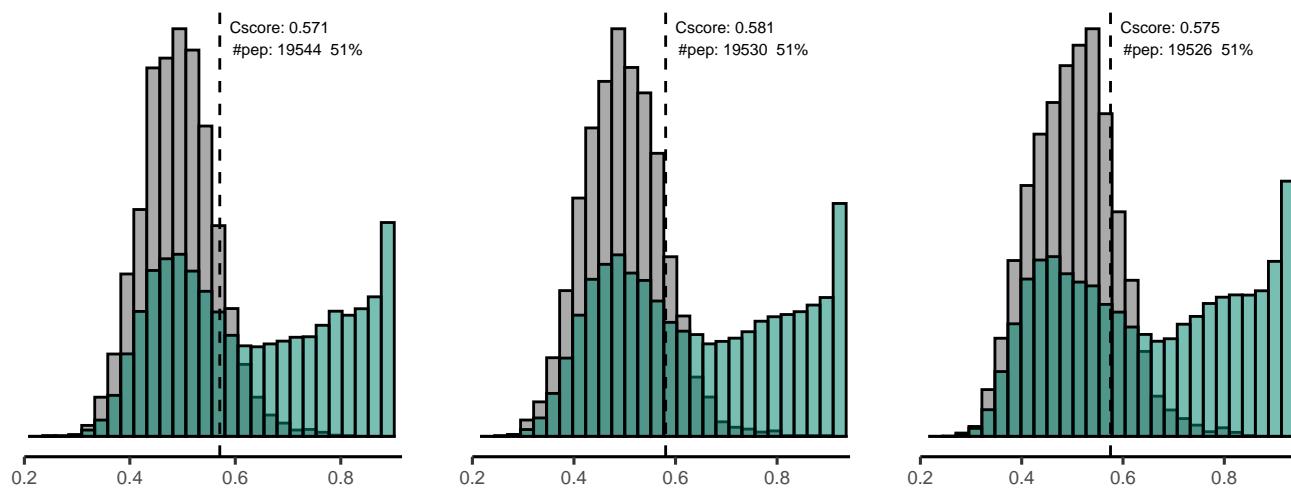
3\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_D920190210\_QX1\_PhGe\_SA\_EASY12-10\_Majken\_B12\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_B3



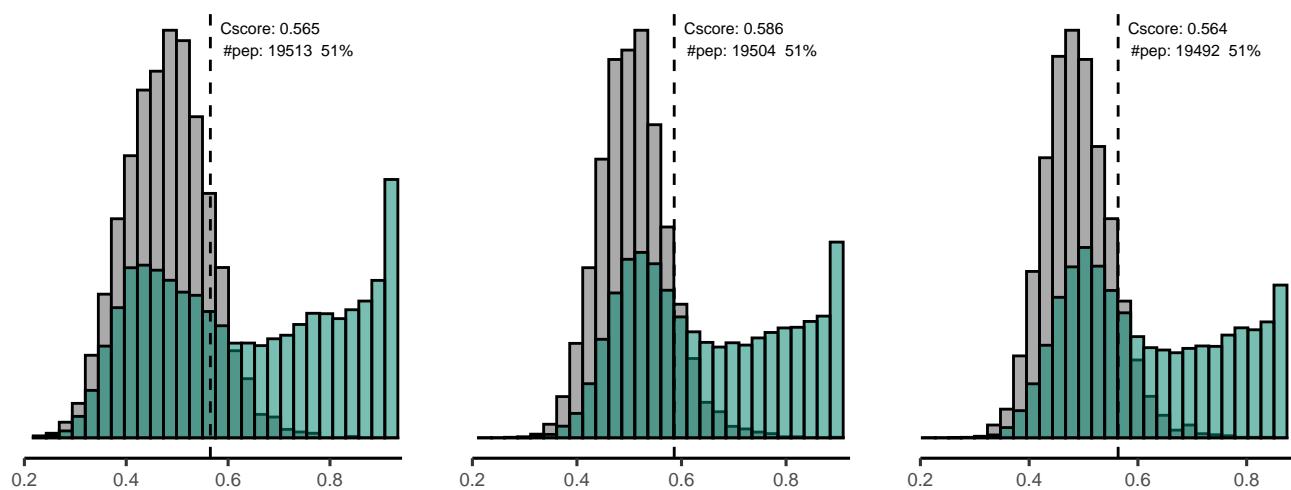
18\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1x20190228\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_E5\_20190210\_QX1\_PhGe\_SA\_EASY12-10\_Majken\_C1\_CSF



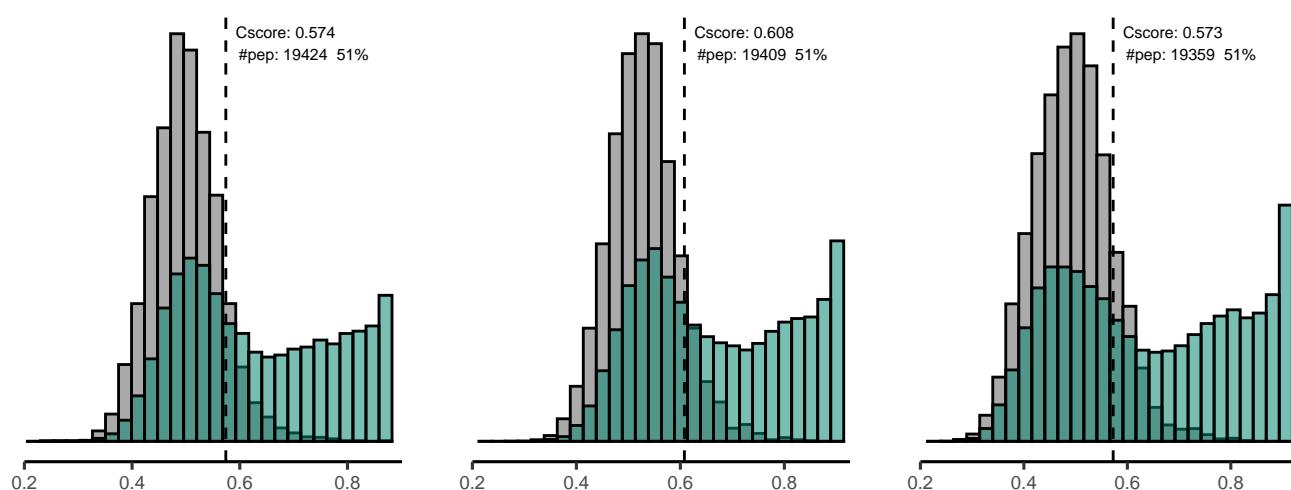
228\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_F20190210\_QX1\_PhGe\_SA\_EASY12-10\_20190210\_QX1\_PhGe\_SA\_EASY12-10\_Majken\_A7\_CSF\_531-12\_c



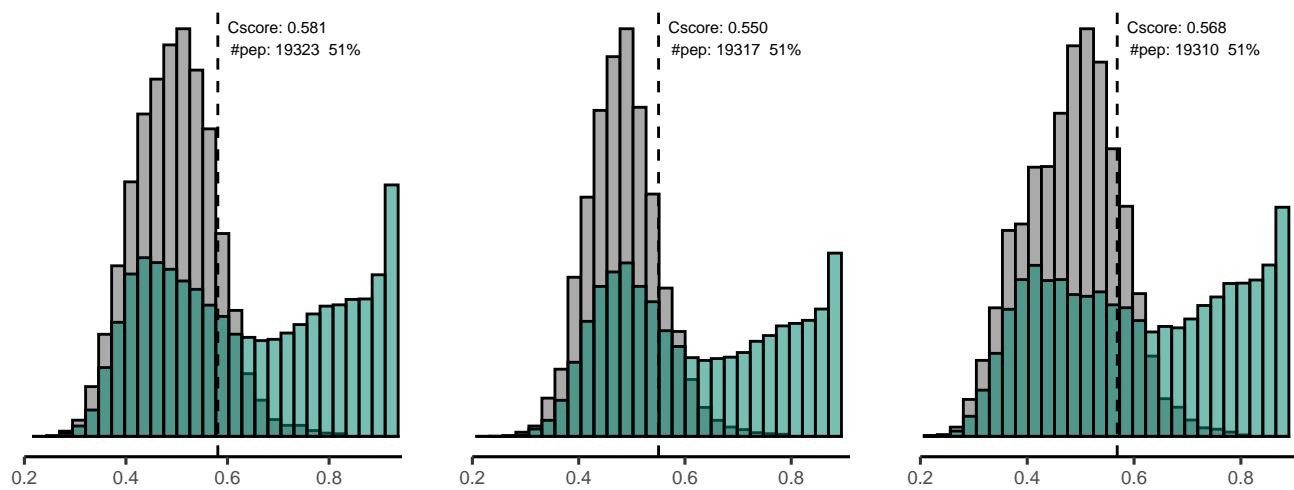
190210\_QX1\_PhGe\_SA\_EASY12-10\_Majken\_20190210\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD120190618\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD1xS1fM1.



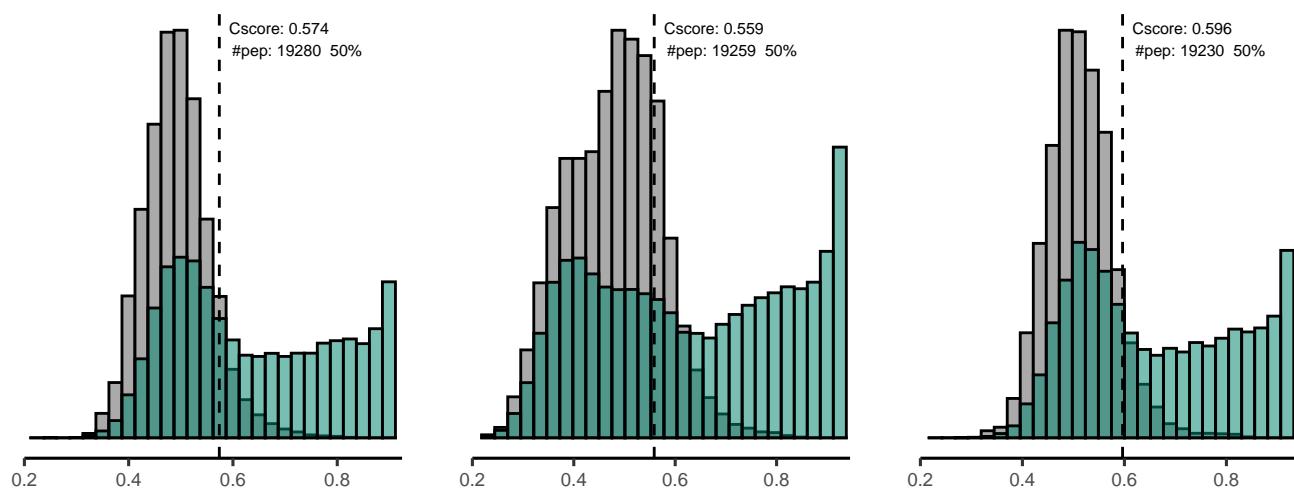
18\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD120190618\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD1xS1fM1\_20190218\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_F8\_190:



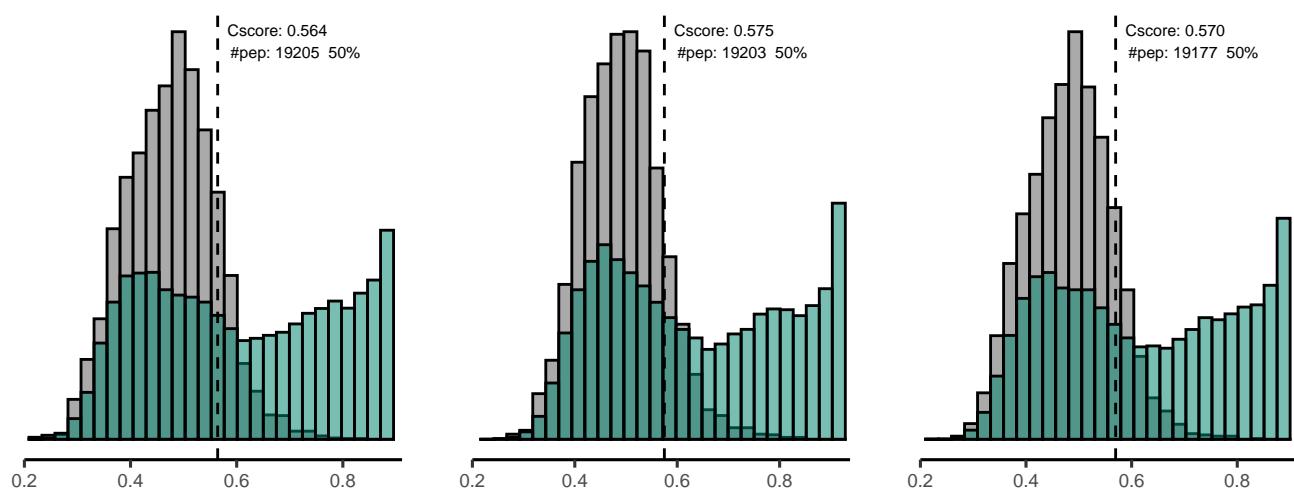
j0210\_QX1\_PhGe\_SA\_EASY12-10\_Majken\_C12\_CSF



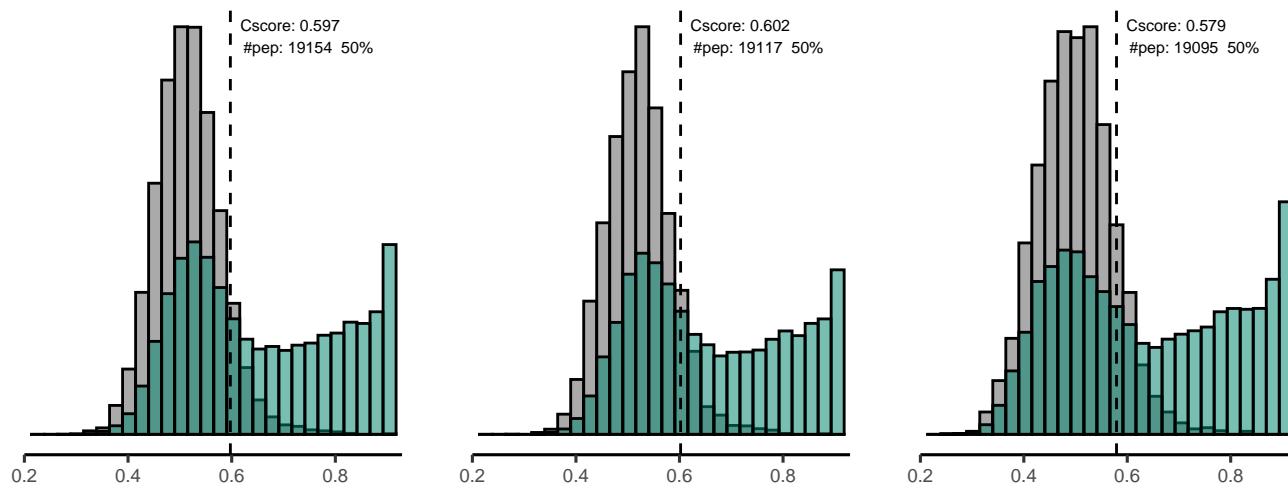
18\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD1xSA0M190210\_QX1\_PhGe\_SA\_EASY12-10\_Majken20880618\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD1xS1fM1\_;



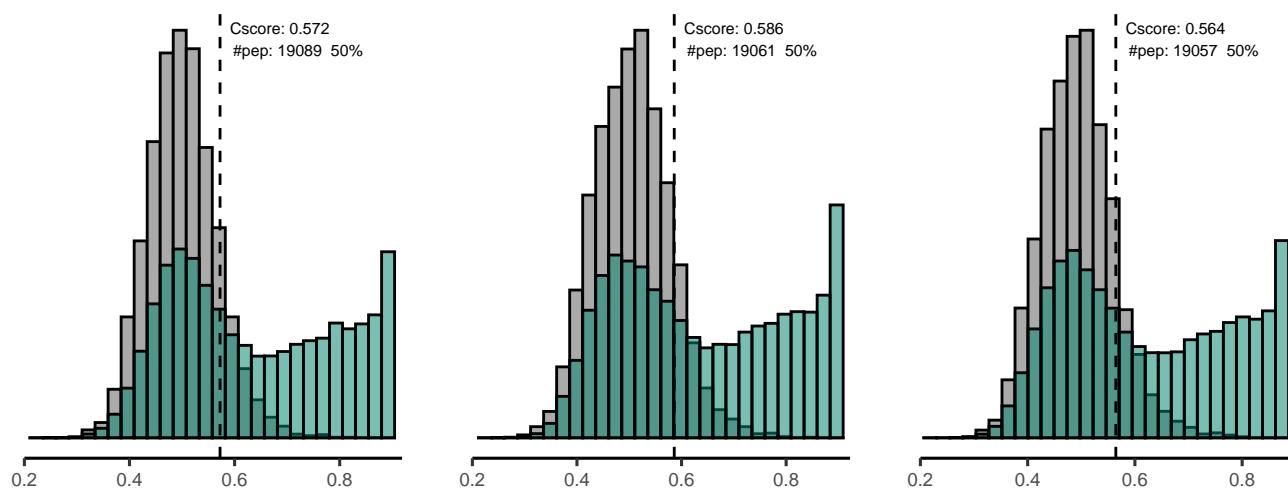
90210\_QX1\_PhGe\_SA\_EASY12-10\_Majken\_C620SF90210\_QX1\_PhGe\_SA\_EASY12-10\_Majken\_B1120S90210\_QX1\_PhGe\_SA\_EASY12-10\_Majken\_B6\_CSF



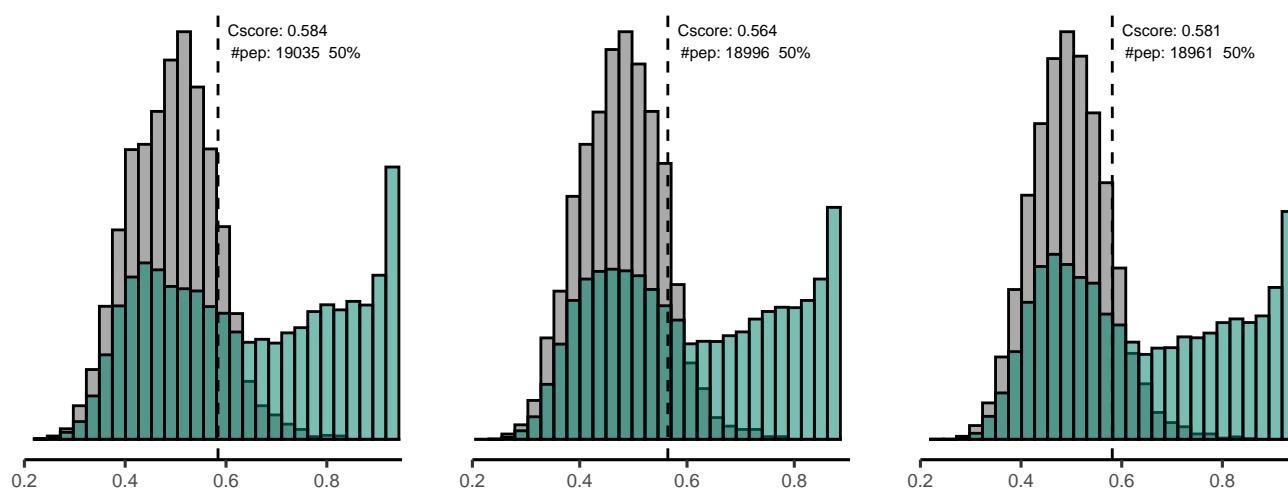
3\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD1x**20100618**\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD1x**2010228**\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_A11\_



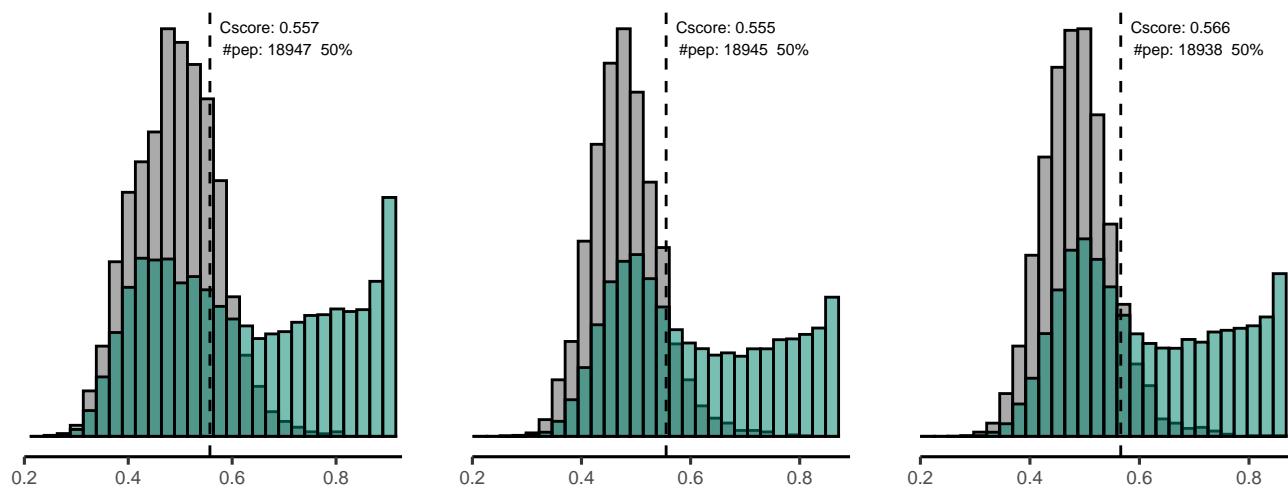
?28\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch**2010228**\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_A**2010228**\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_H9\_



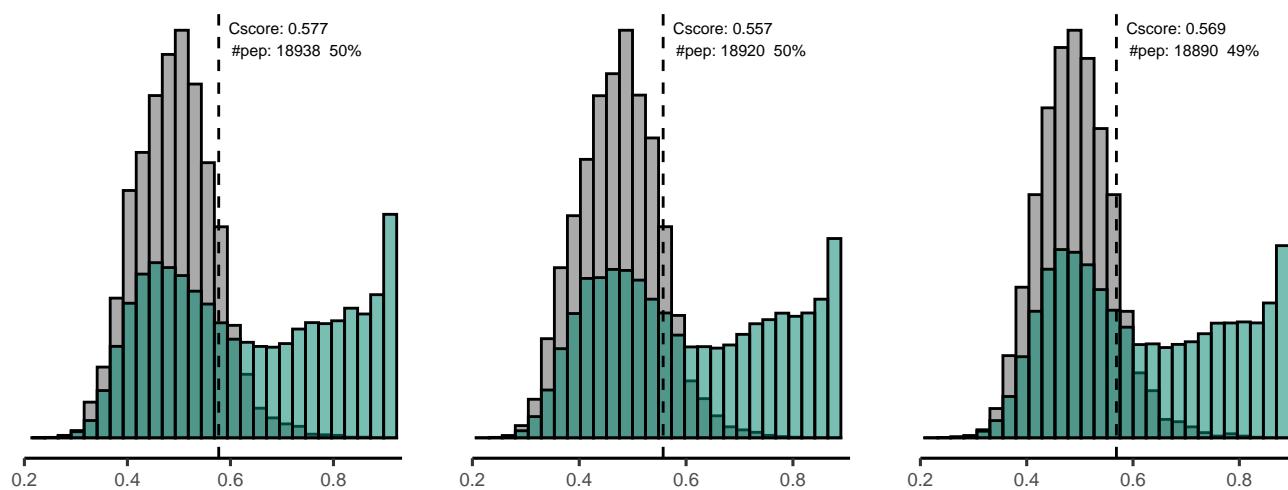
228\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch**2010228**\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch**2010210**\_QX1\_PhGe\_SA\_EASY12-10\_Majken\_B2\_CSF



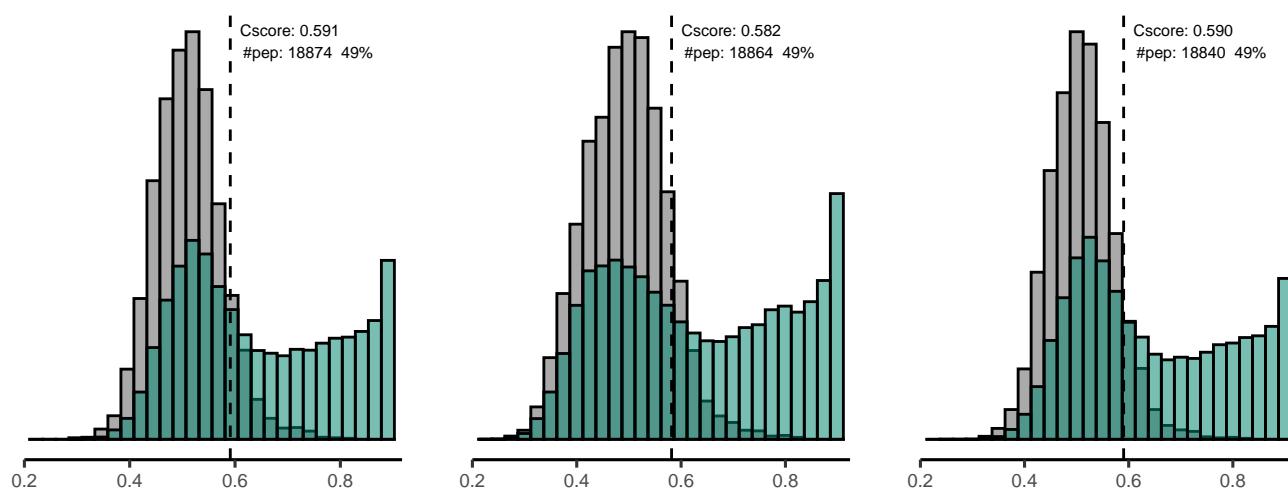
!28\_QX1\_PhGe\_SA\_EASY12-10\_Deigendes20100618\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD1xS1fm1.



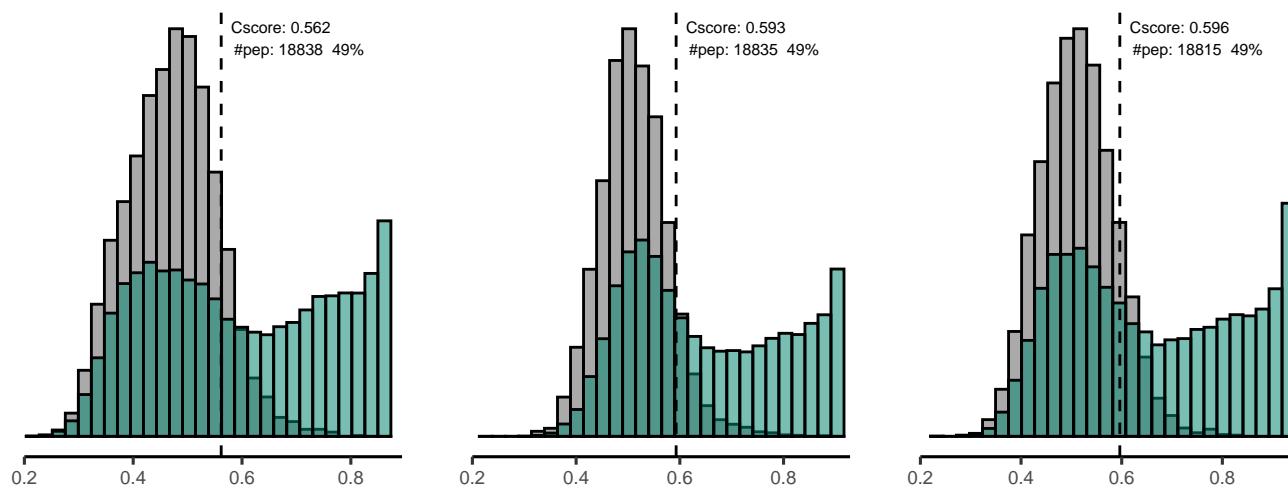
\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_D2010228\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch2010228\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_G12\_



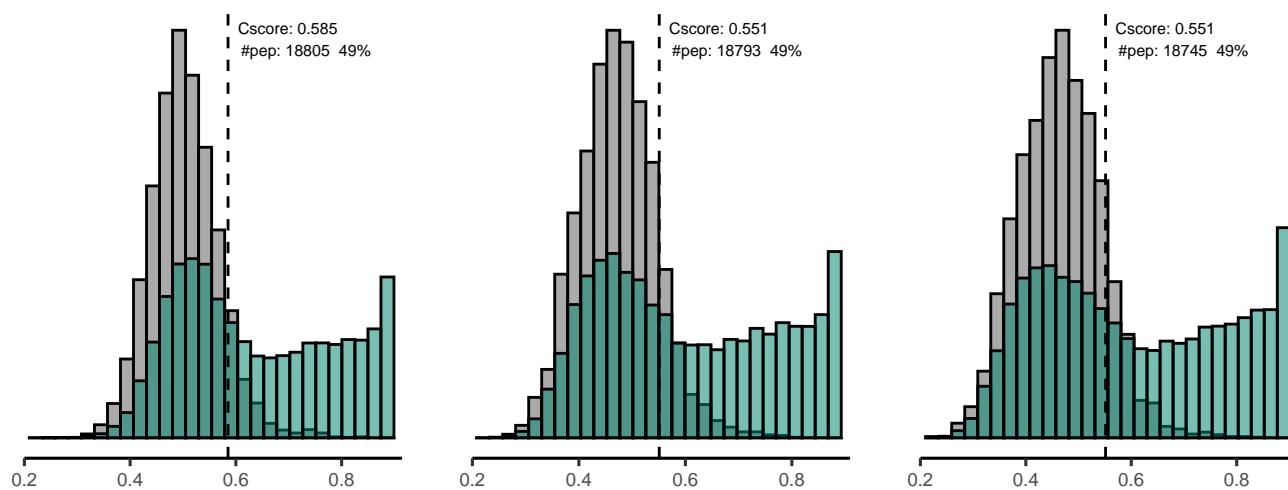
18\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD1xS1fm1\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch20100618\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD1xS1fm1.



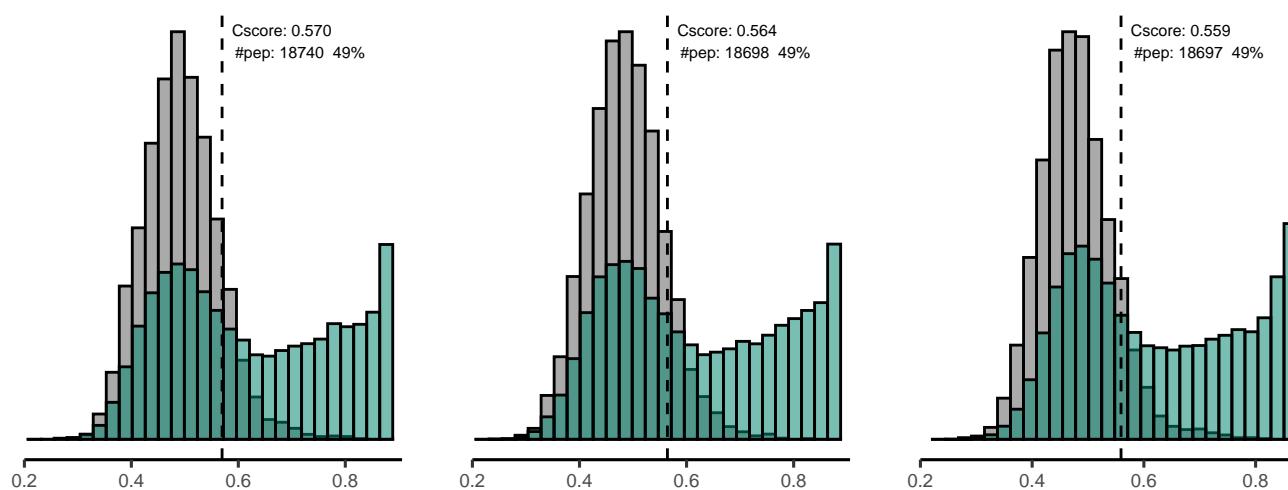
0210\_QX1\_PhGe\_SA\_EASY12-10\_Majken\_C6168718\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD1xS1fm1 0228\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_B2\_



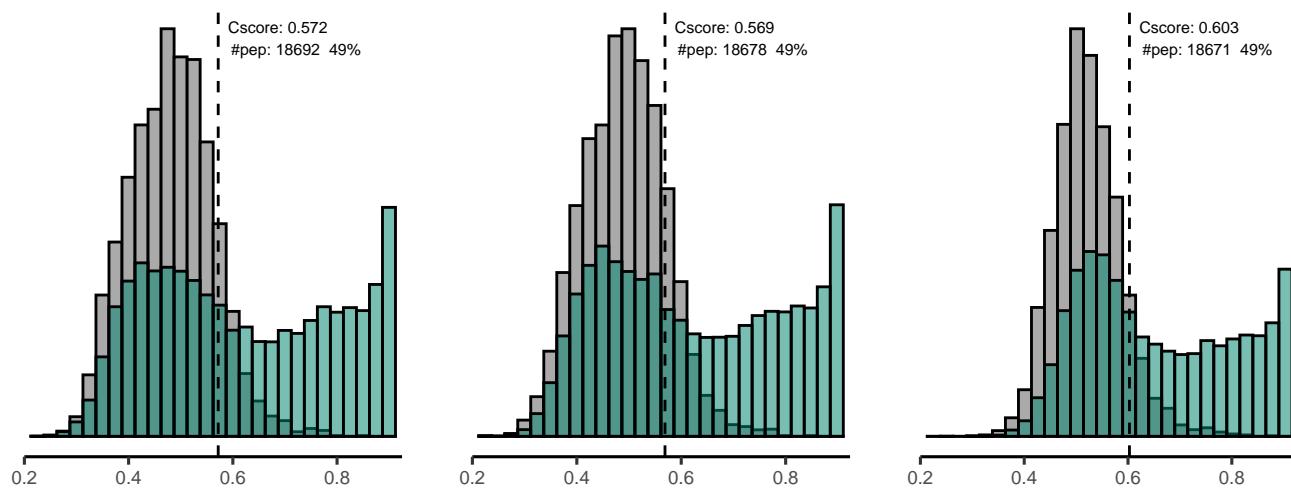
18\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD1xS1fm1 0228\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_A7\_20190210\_QX1\_PhGe\_SA\_EASY12-10\_Majken\_E12\_C



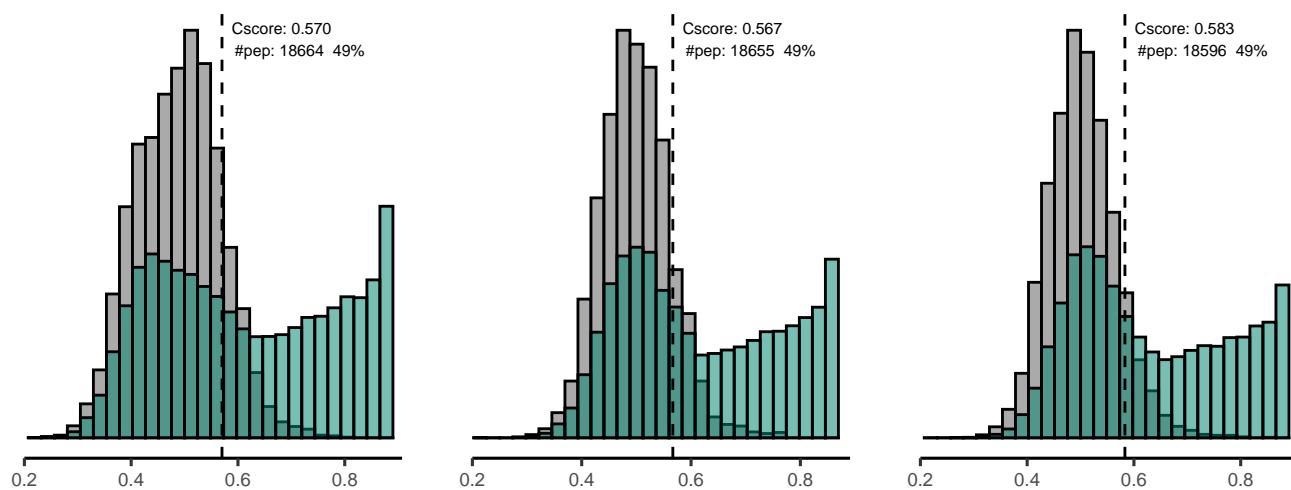
228\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_A4 20190210\_QX1\_PhGe\_SA\_EASY12-10\_Majken\_C6168718\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD1xS1fm1.



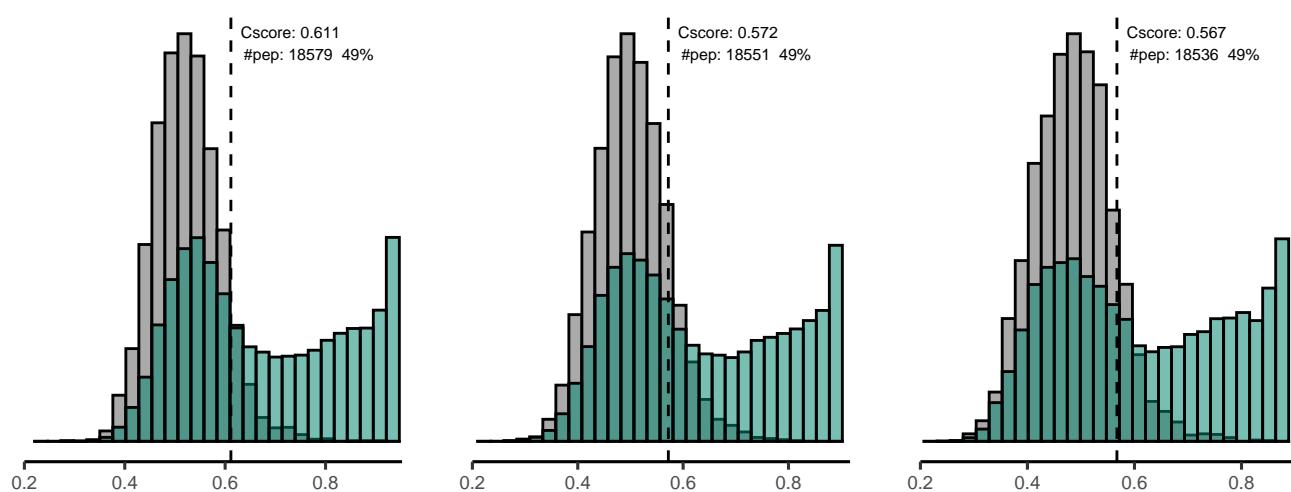
1228\_QX1\_PhGe\_SA\_EASY12-10\_Deigendes20106228\_QX1\_PhGe\_SA\_EASY12-10\_Deigendes2010618\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD1xS1fM1.



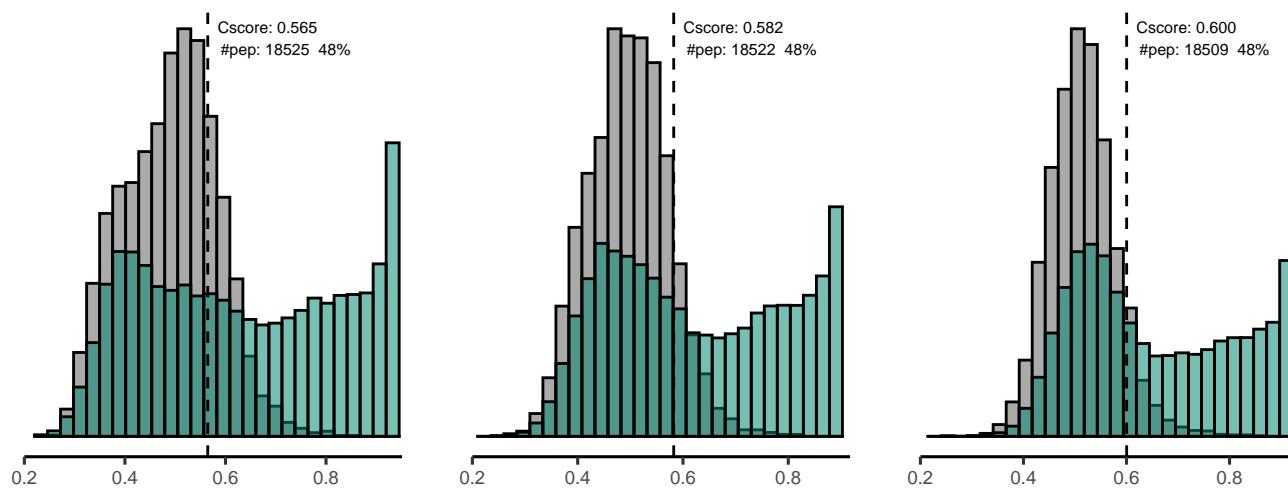
1228\_QX1\_PhGe\_SA\_EASY12-10\_Deigendes20106228\_QX1\_PhGe\_SA\_EASY12-10\_Deigendes2010618\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD1xS1fM1.



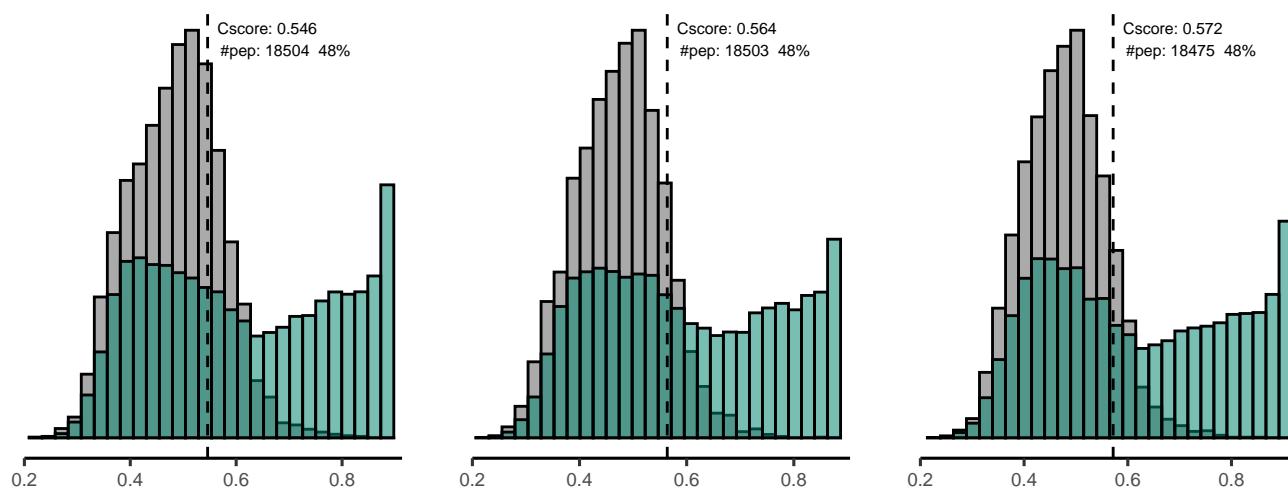
18\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD1xS1fM1 20106228\_QX1\_PhGe\_SA\_EASY12-10\_Deigendes2010618\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD1xS1fM1.



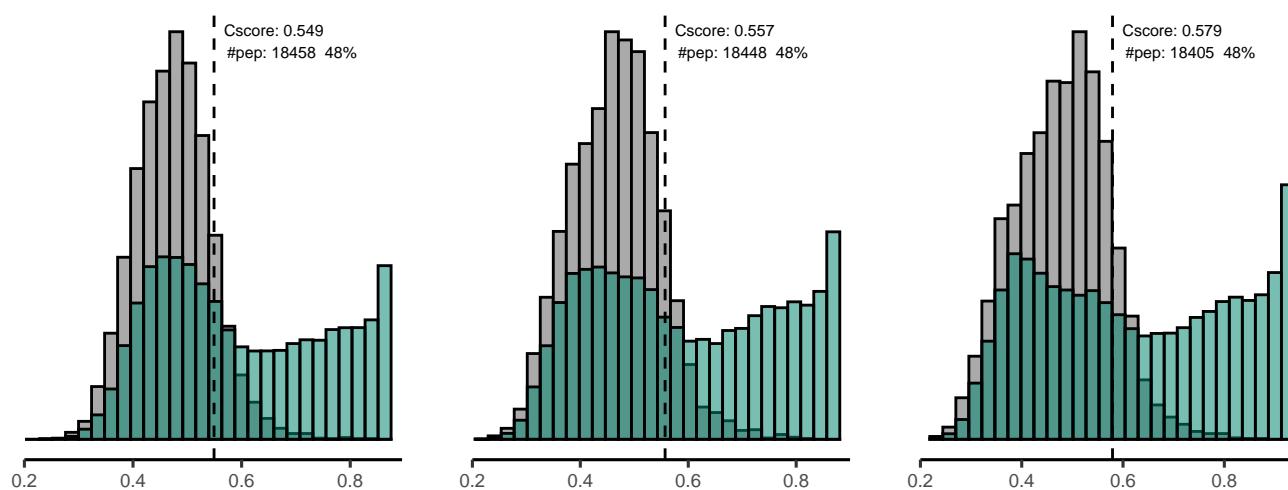
0210\_QX1\_PhGe\_SA\_EASY12-10\_Majken\_B9\_CSF



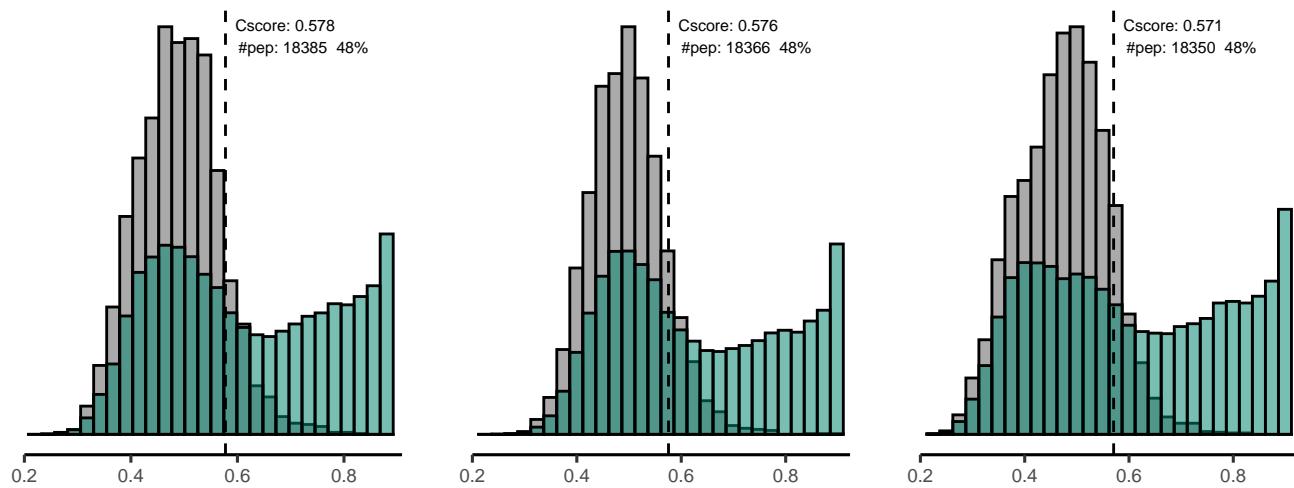
)228\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_E11\_828720390210\_QX1\_PhGe\_SA\_EASY12-10\_Majken\_B9\_CSF



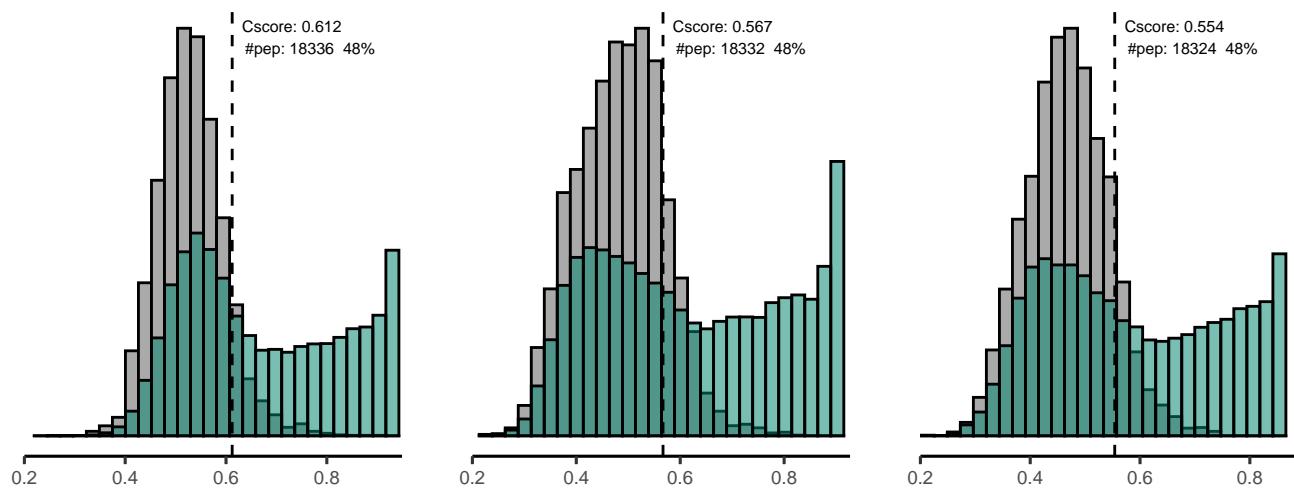
228\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_E320290210\_QX1\_PhGe\_SA\_EASY12-10\_Majken\_C7\_CSF



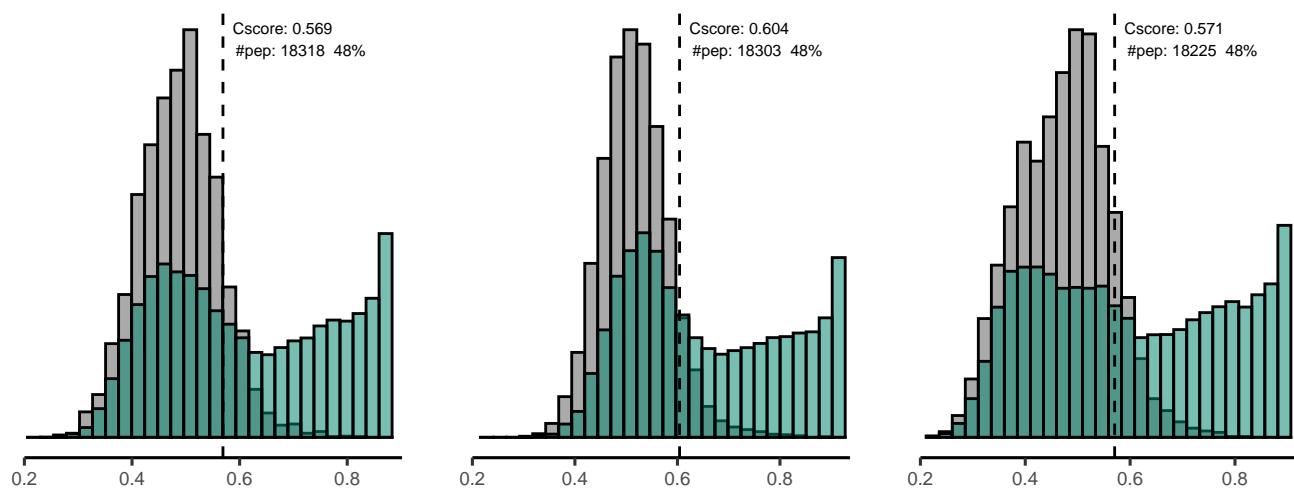
190210\_QX1\_PhGe\_SA\_EASY12-10\_Majken\_E5\_C0190210\_QX1\_PhGe\_SA\_EASY12-10\_Majken\_E20C0210\_QX1\_PhGe\_SA\_EASY12-10\_Majken\_D4\_CSF



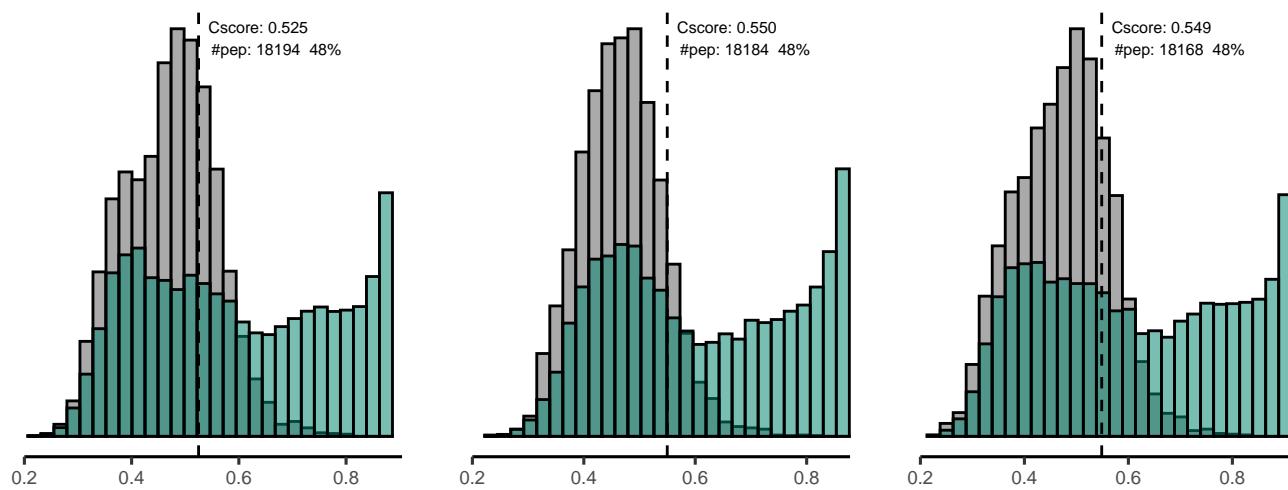
18\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD1xS20M0228\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch20390210\_QX1\_PhGe\_SA\_EASY12-10\_Majken\_B3\_CSF



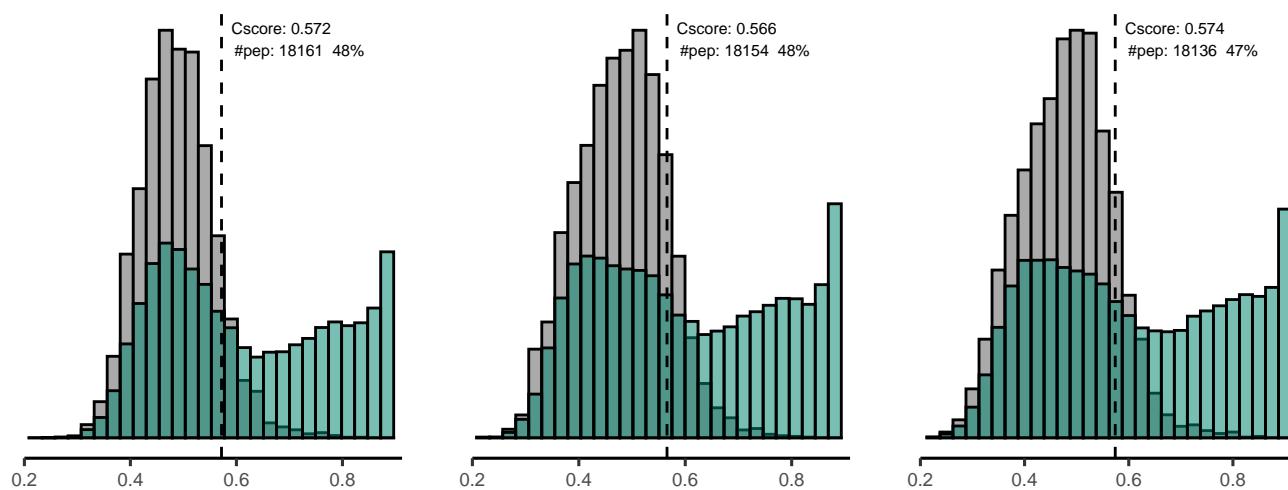
228\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch2016618\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD1xS21M90210\_QX1\_PhGe\_SA\_EASY12-10\_Majken\_D5\_CSF



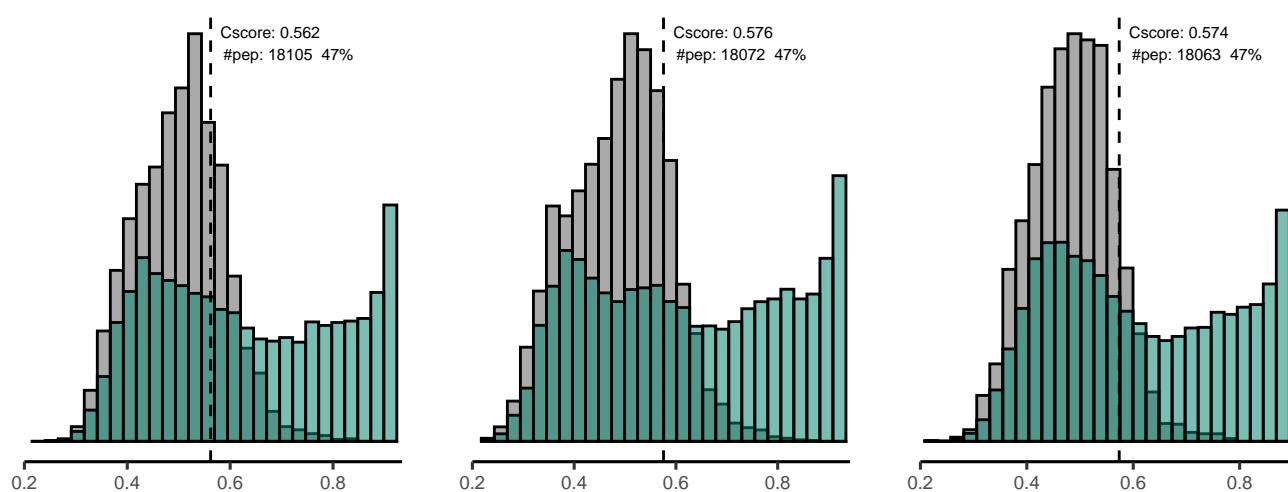
228\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_20190210\_QX1\_PhGe\_SA\_EASY12-10\_Majken\_A1020190210\_QX1\_PhGe\_SA\_EASY12-10\_Majken\_C9\_CSF



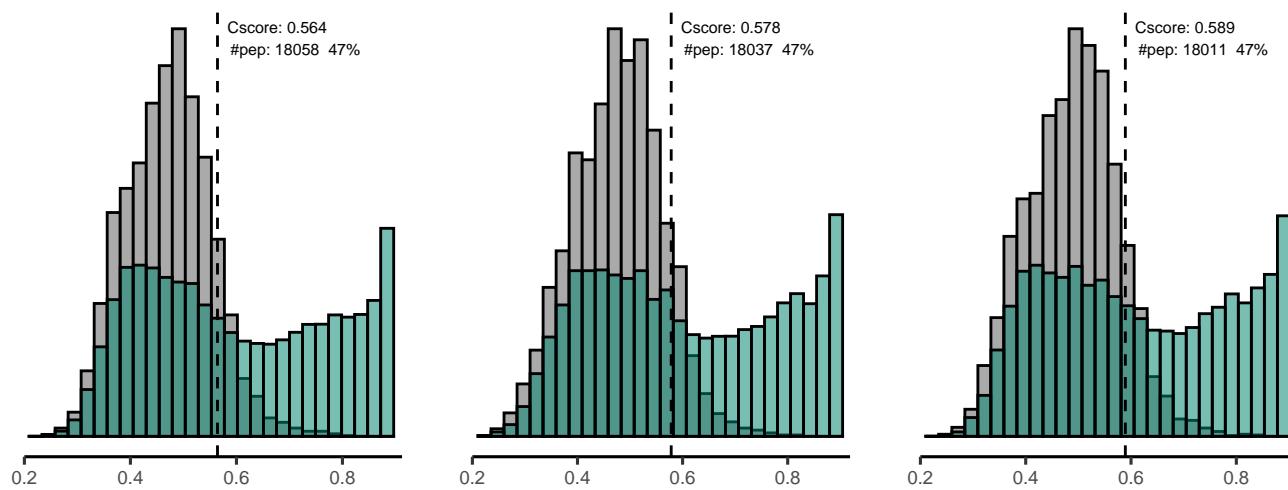
\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_F20190228\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_G20190210\_QX1\_PhGe\_SA\_EASY12-10\_Majken\_F8\_CS



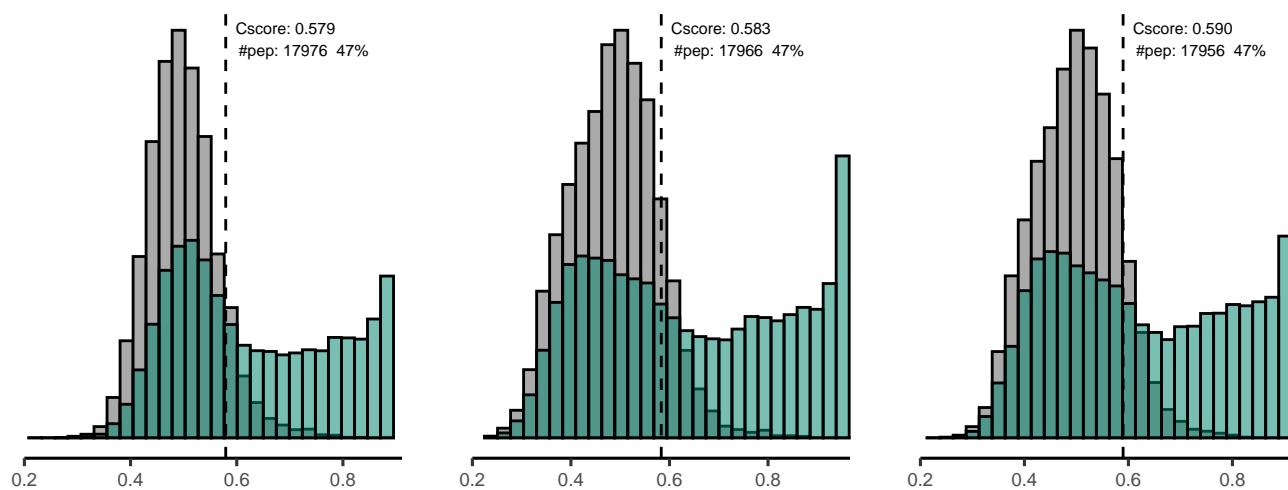
)228\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_20190210\_QX1\_PhGe\_SA\_EASY12-10\_Majken\_20190224\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_E8\_1291885



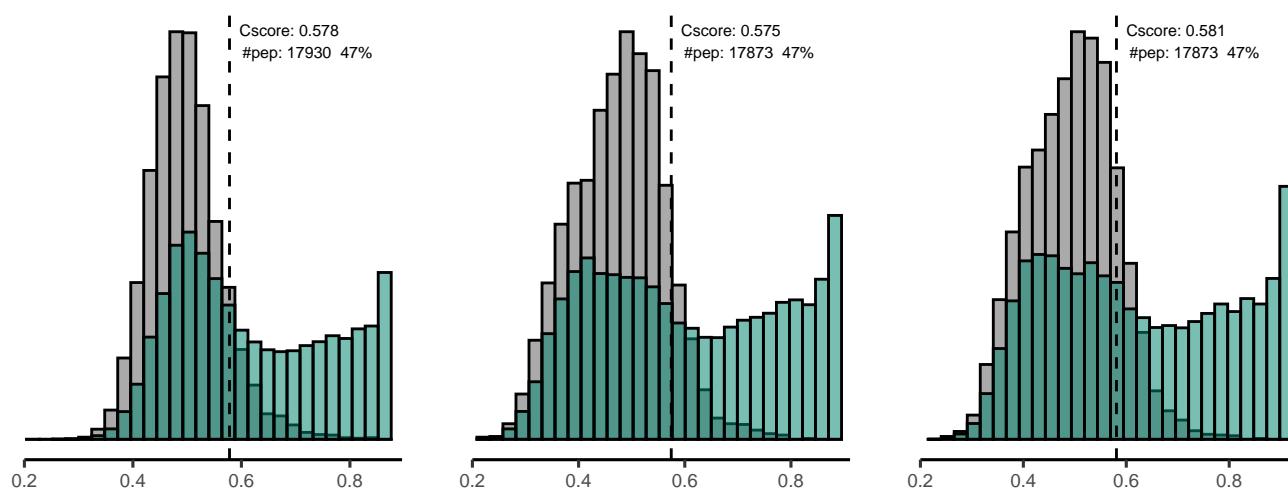
J228\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_A@20190210\_QX1\_PhGe\_SA\_EASY12-10\_Majken\_F4\_20190210\_QX1\_PhGe\_SA\_EASY12-10\_Majken\_E2\_C5



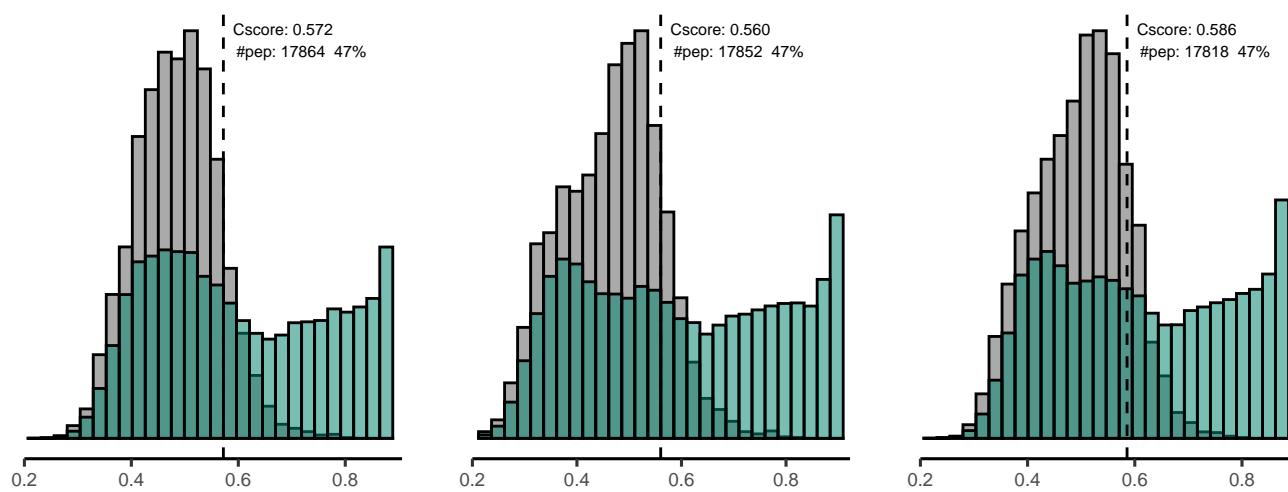
18\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD1xS1D1@20190210\_QX1\_PhGe\_SA\_EASY12-10\_Majken\_A12@20190228\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_B5\_



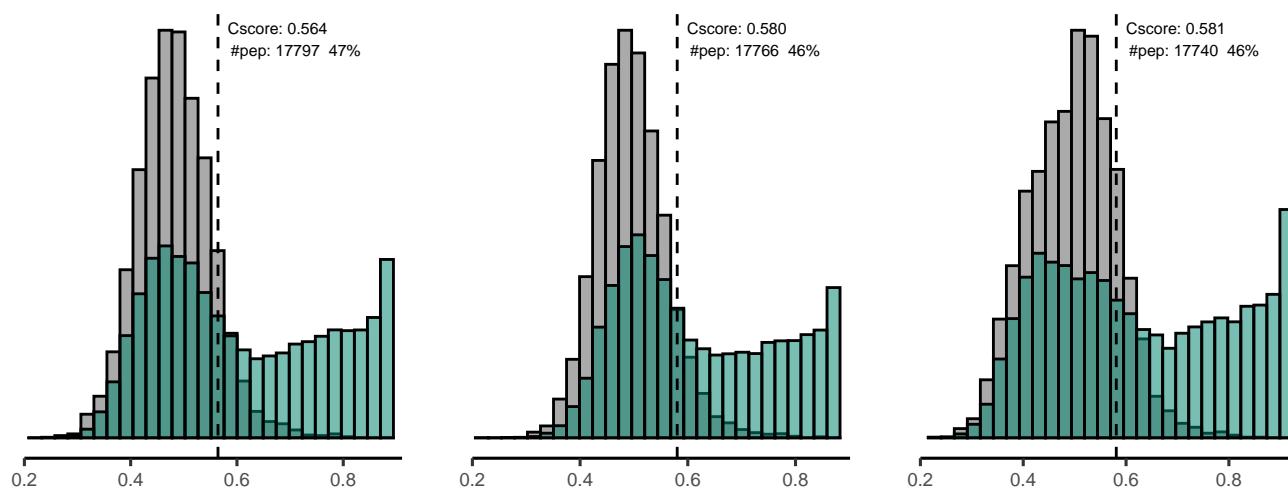
18\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD1xS1D1@20190210\_QX1\_PhGe\_SA\_EASY12-10\_Majken\_20190228\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_E9\_828



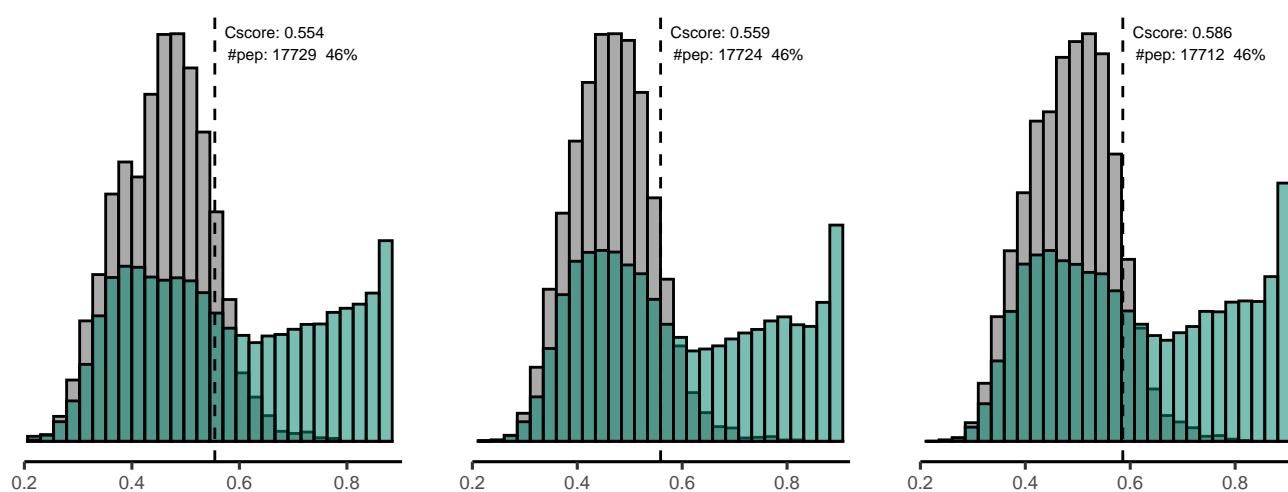
?228\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_B10



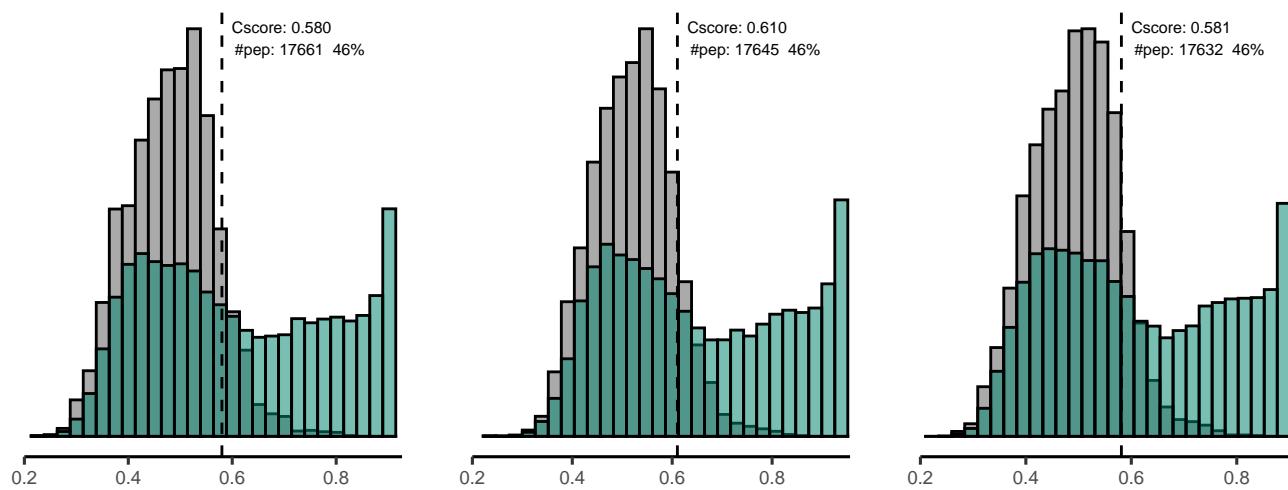
?28\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_B10



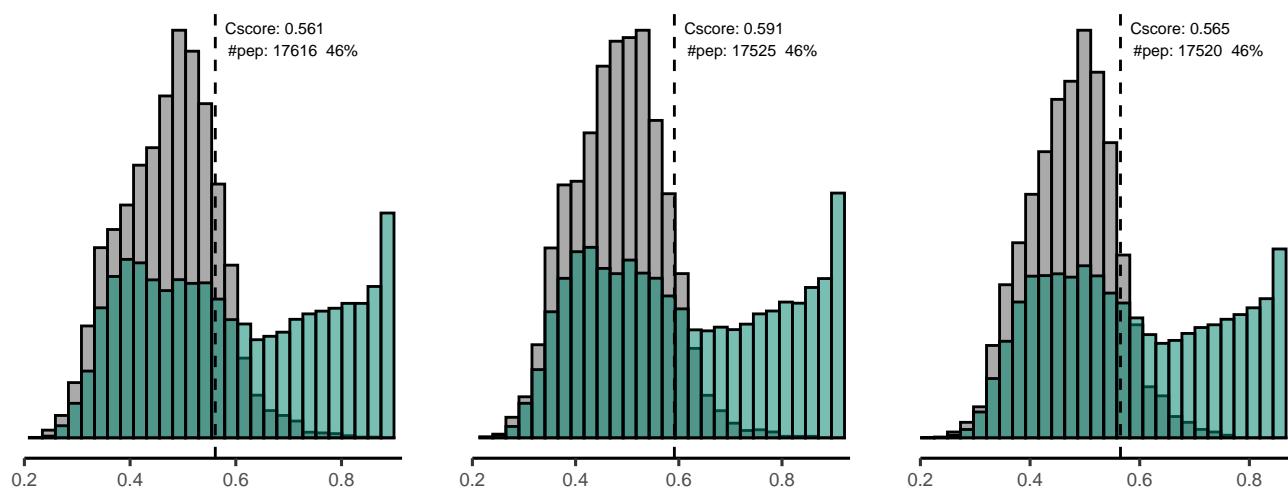
190210\_QX1\_PhGe\_SA\_EASY12-10\_Majken\_B10



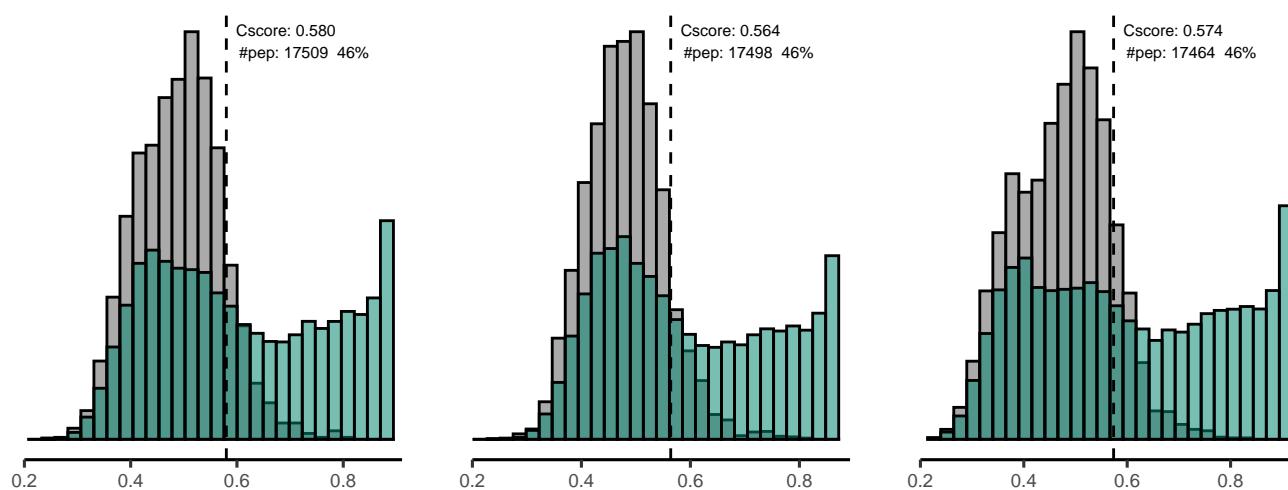
!228\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_B20190210\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_D120190228\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_A2\_



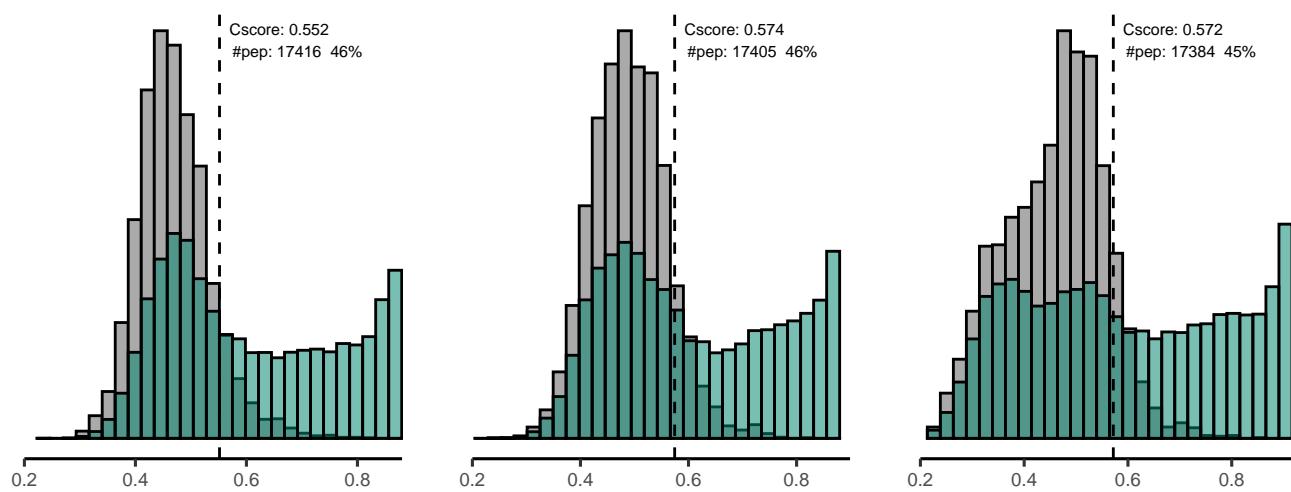
?28\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_B20190210\_QX1\_PhGe\_SA\_EASY12-10\_Majken\_E20190228\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_D1\_



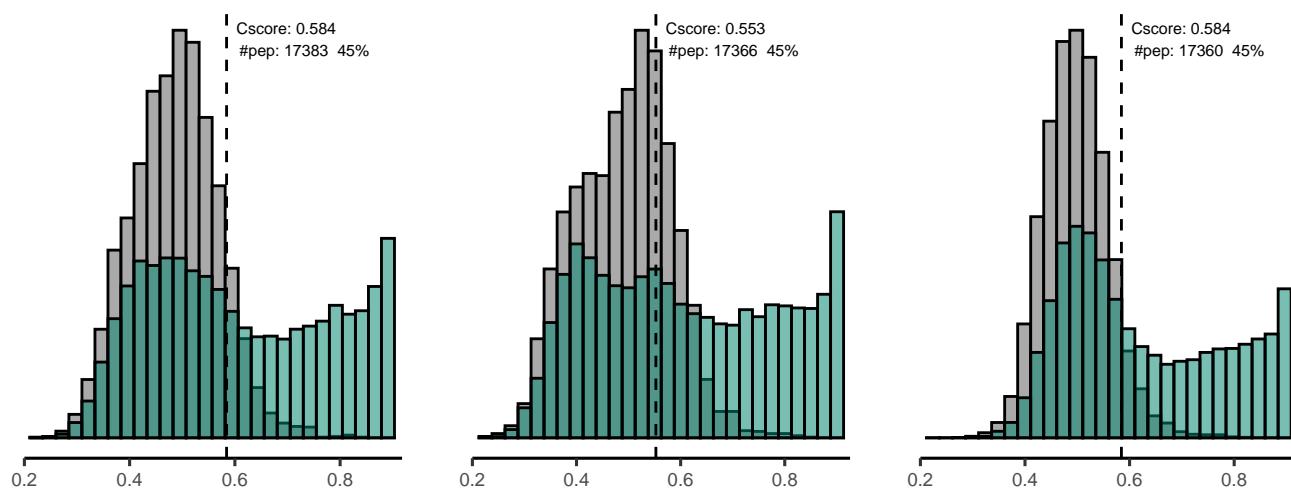
\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_A20190228\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_D20190210\_QX1\_PhGe\_SA\_EASY12-10\_Majken\_A8\_CSF



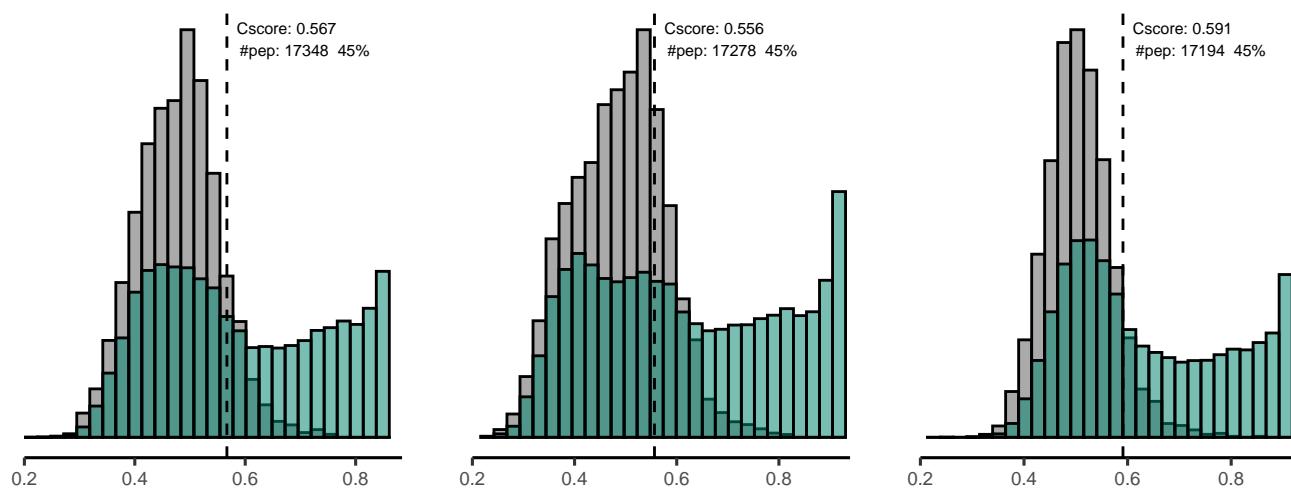
18\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD1xS1fM1 20190228\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_20180618\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD1xS1fM1



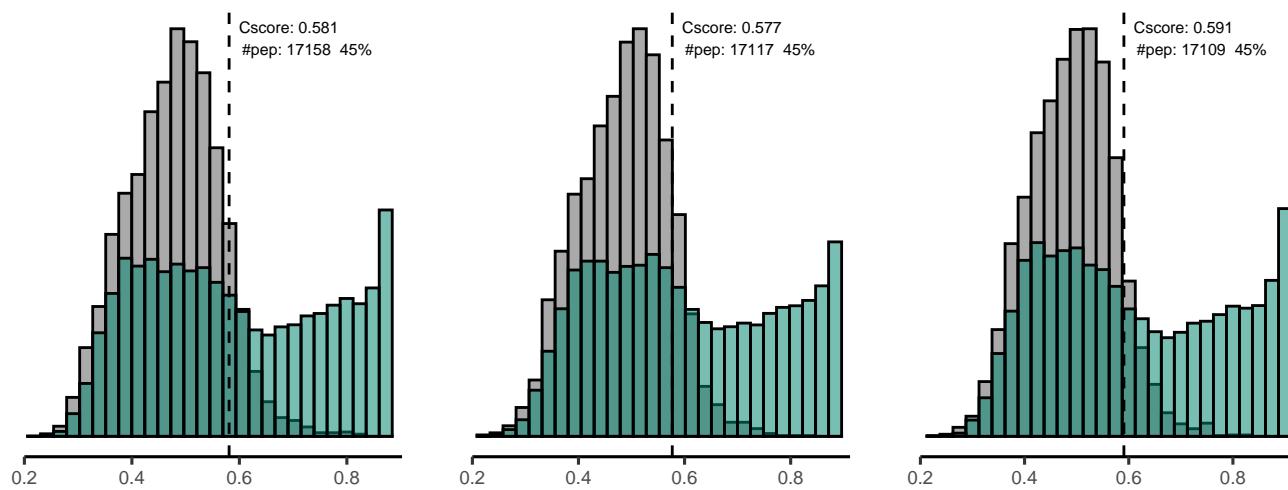
190210\_QX1\_PhGe\_SA\_EASY12-10\_Majken\_20190228\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_20180618\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD1xS1fM1



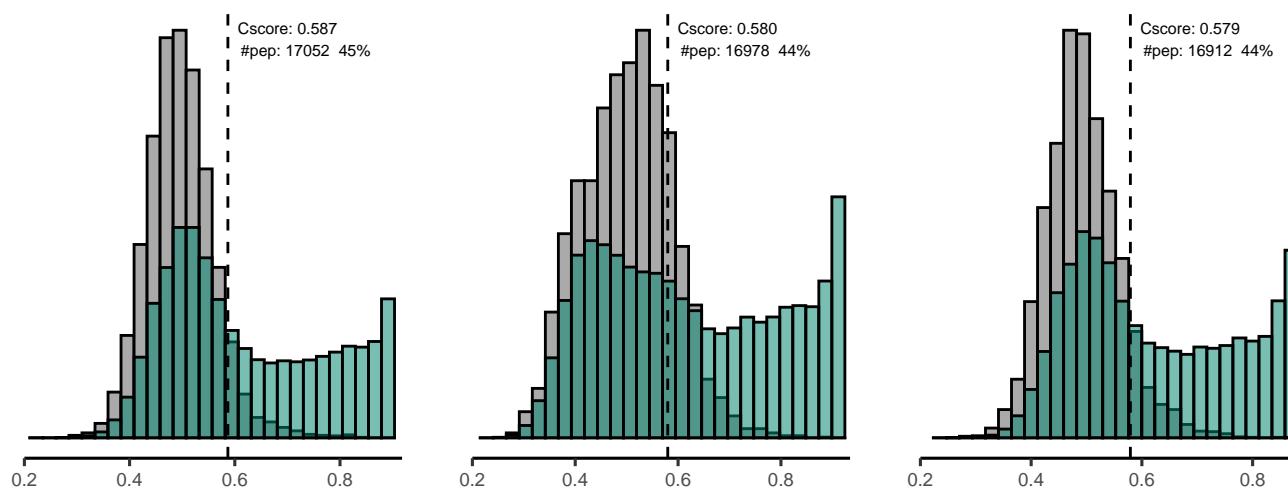
228\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_20190228\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_20180618\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD1xS1fM1



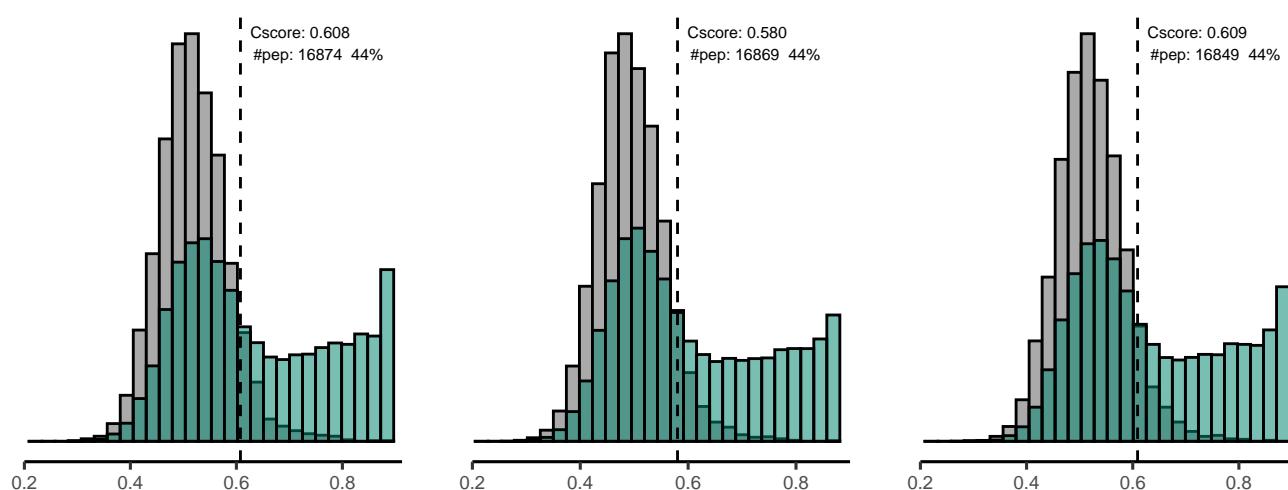
1228\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_B9\_



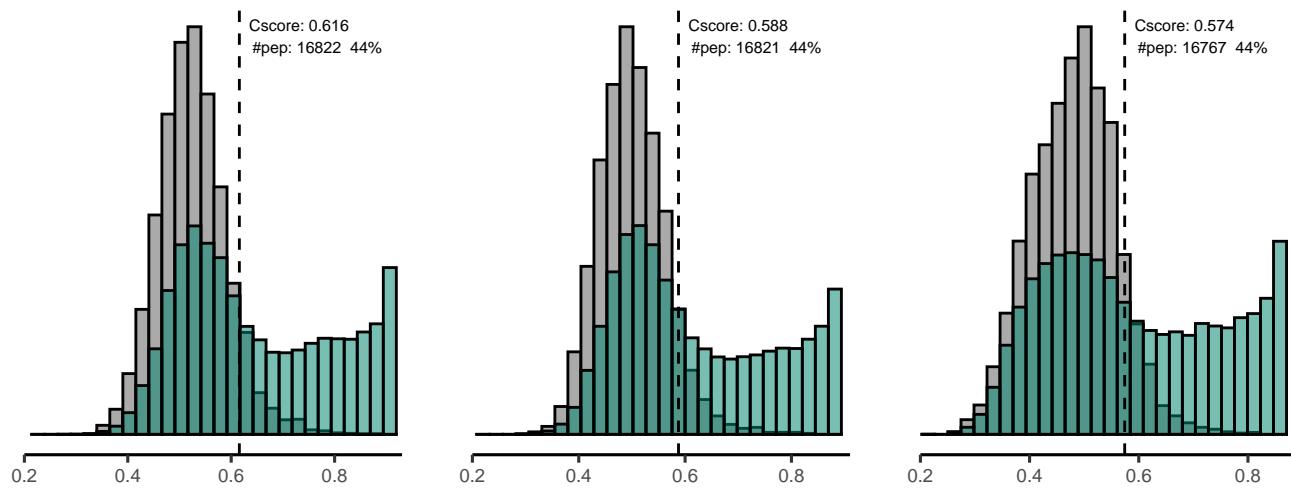
18\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD1xS1fM1\_



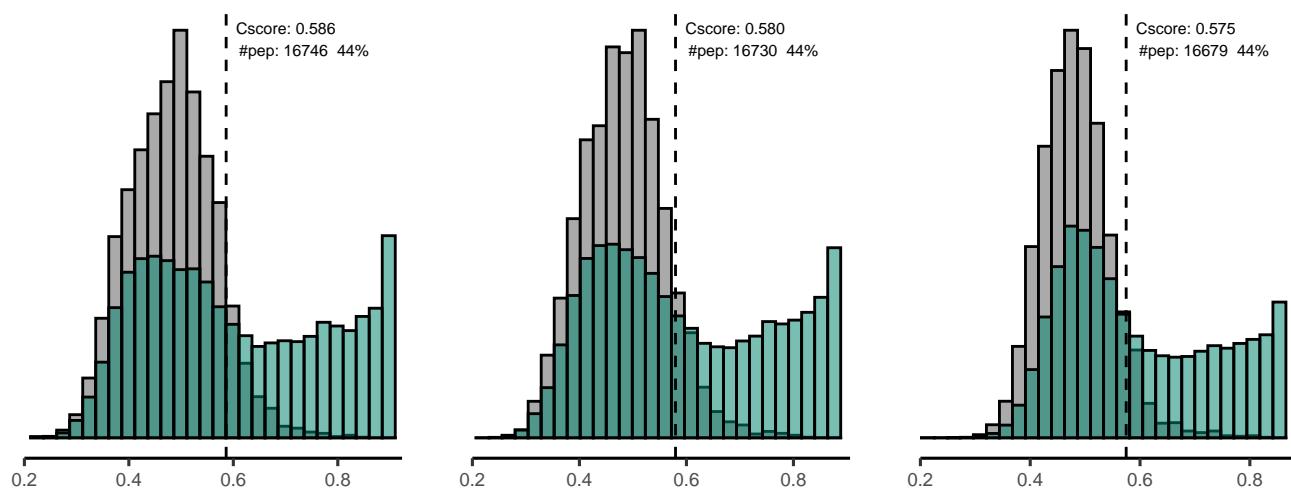
18\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD1xS1fM1\_



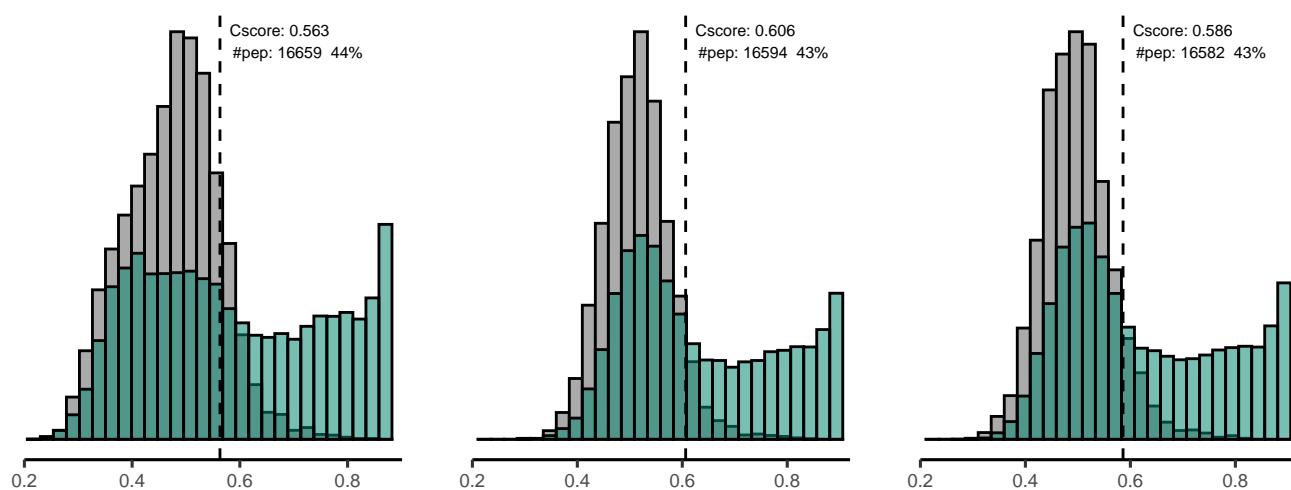
18\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD1xS1fM1 20190728\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD1xS1fM1



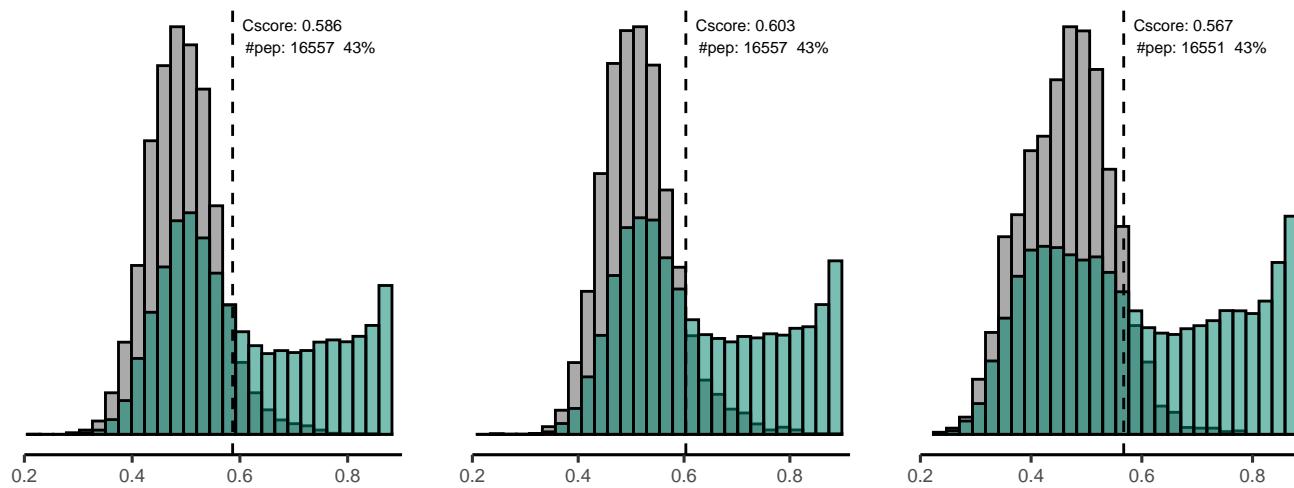
190210\_QX1\_PhGe\_SA\_EASY12-10\_Maj 20190728\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_F20180618\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD1xS1fM1



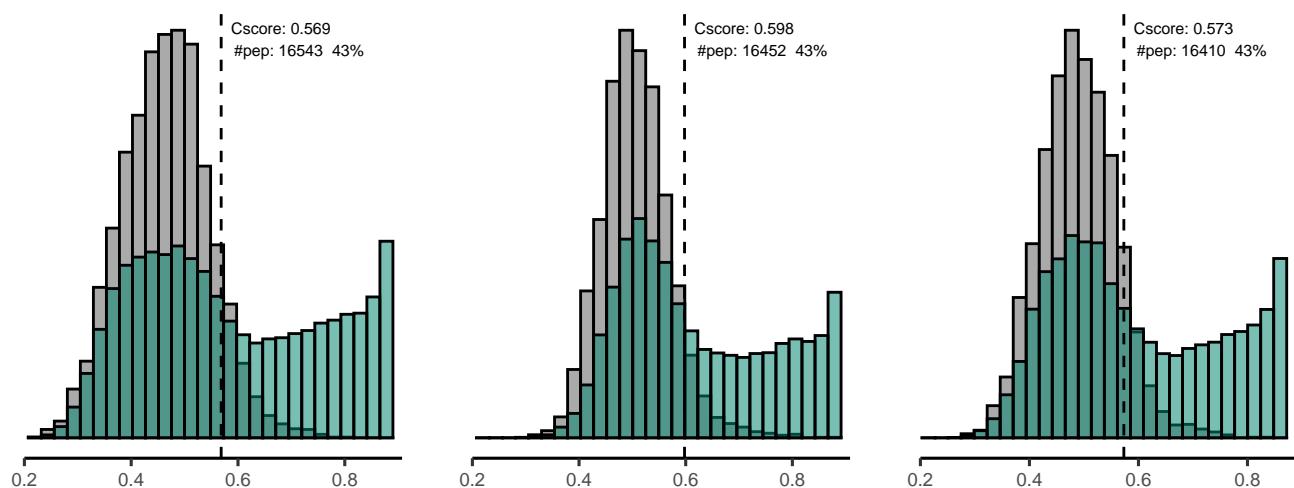
228\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch 20190618\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD1xS1fM1 20190618\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD1xS1fM1



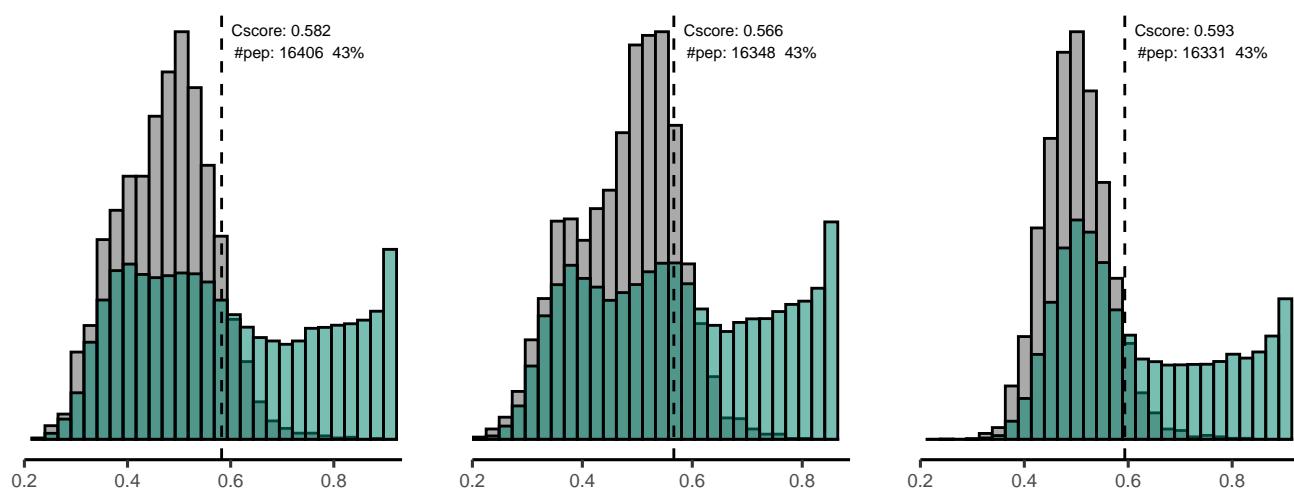
18\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD1xS1fM1 18\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-2xD1xS1fM1 18\_QX0\_JaBa\_SA\_EASY12-10\_Deigendesch\_F9\_828801



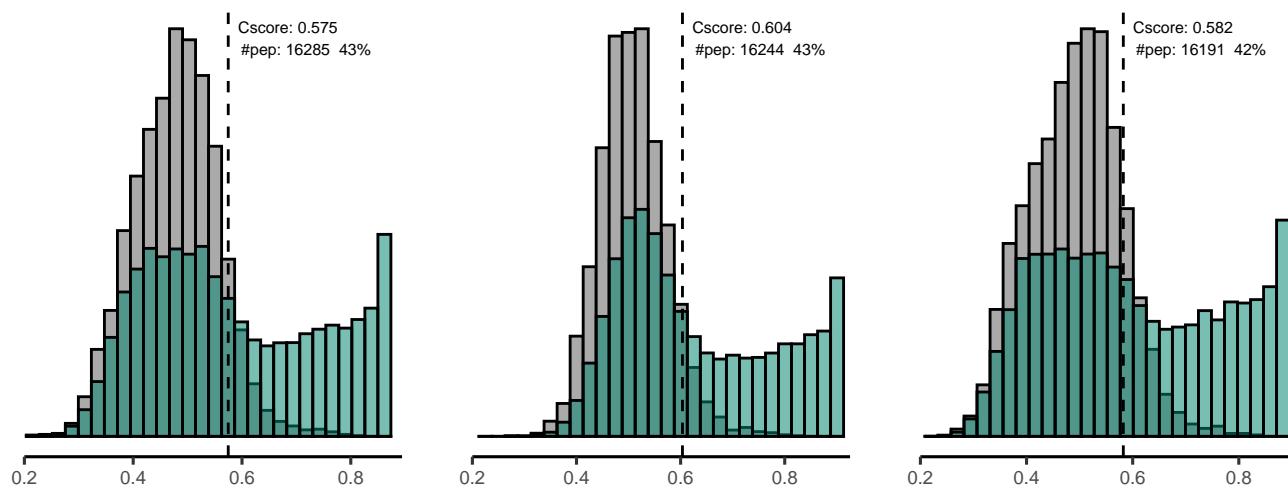
190210\_QX1\_PhGe\_SA\_EASY12-10\_Majken 2018\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD1xS1fM1 2018\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_C5\_



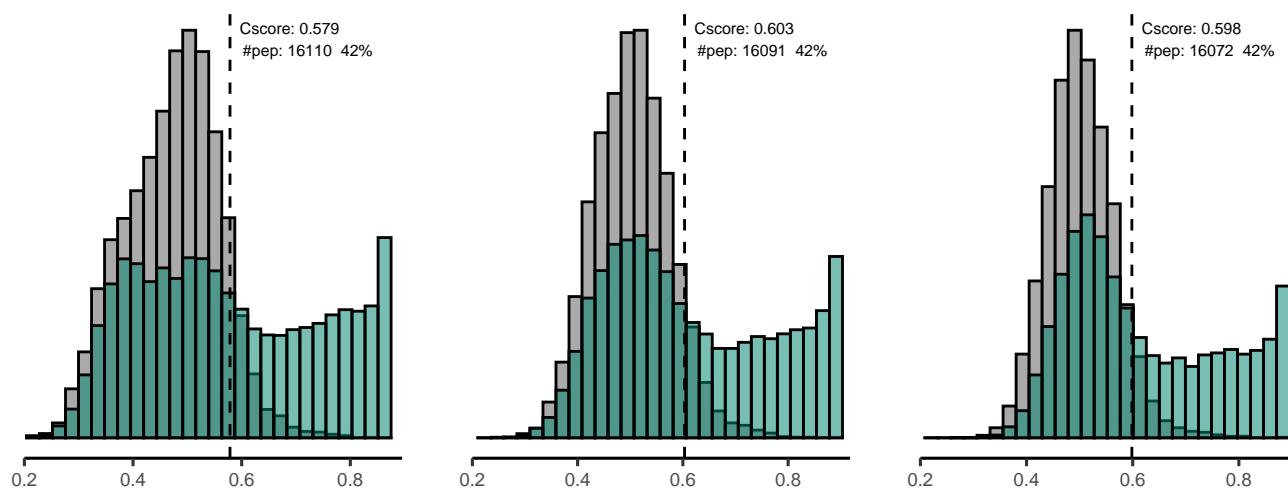
190210\_QX1\_PhGe\_SA\_EASY12-10\_Majken 2018\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD1xS1fM1 2018\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD1xS1fM1



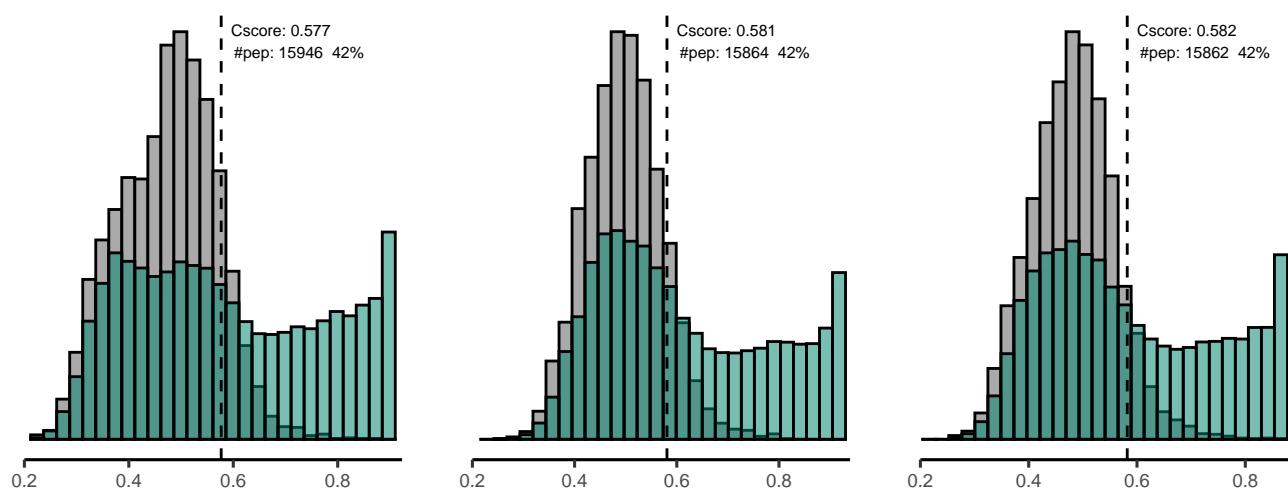
\\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_2018\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8201x01281\QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_F6\_1291899



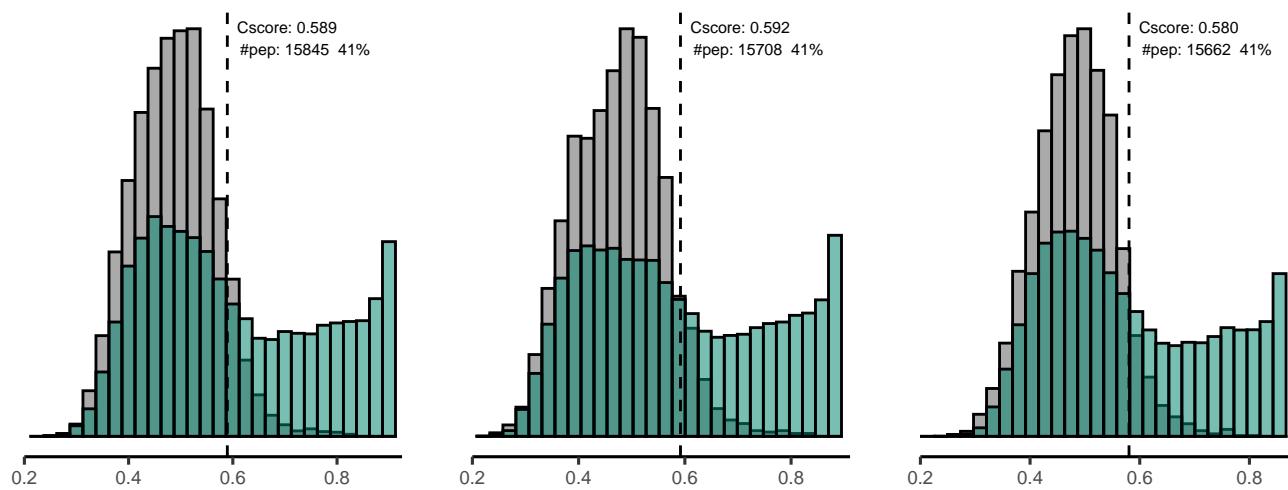
\\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_20190228\QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_20190618\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD1xS1fM1.



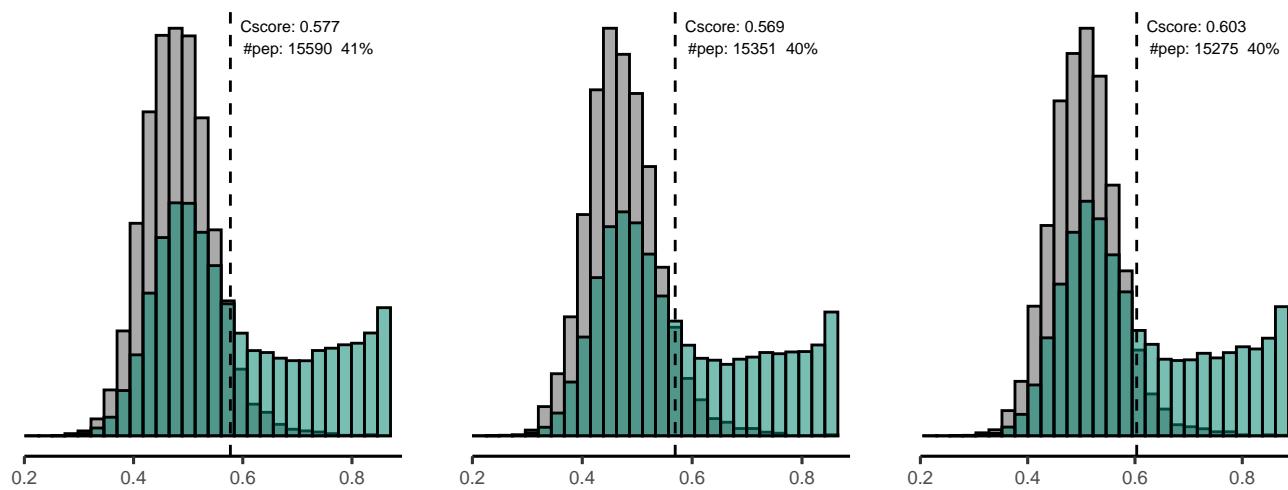
190210\_QX1\_PhGe\_SA\_EASY12-10\_Maj\_20190228\QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_E\190228\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_A12\_



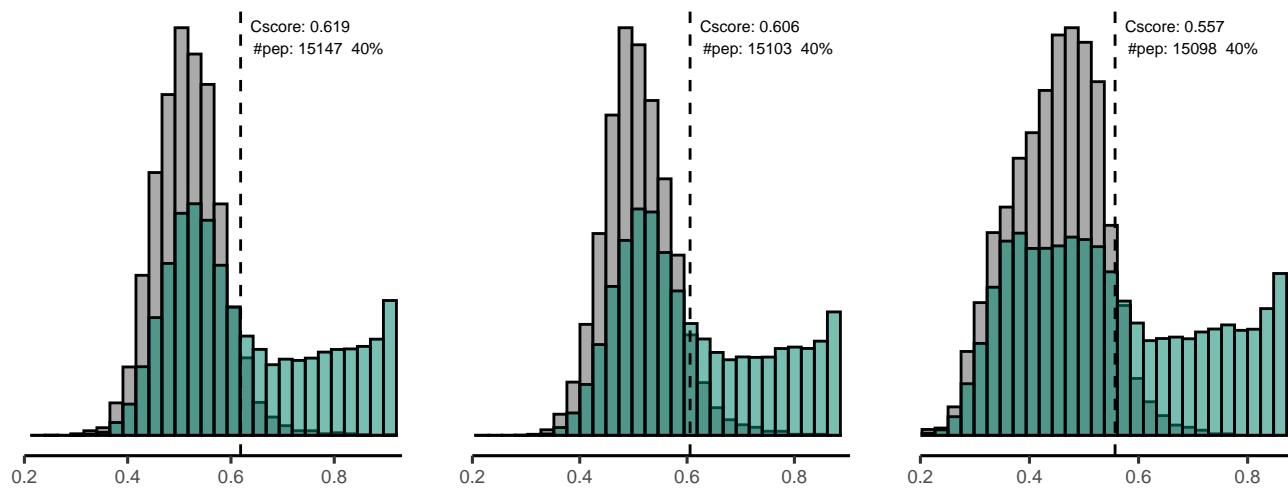
1228\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_D20190210\_QX1\_PhGe\_SA\_EASY12-10\_Majken\_20190228\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_E10\_



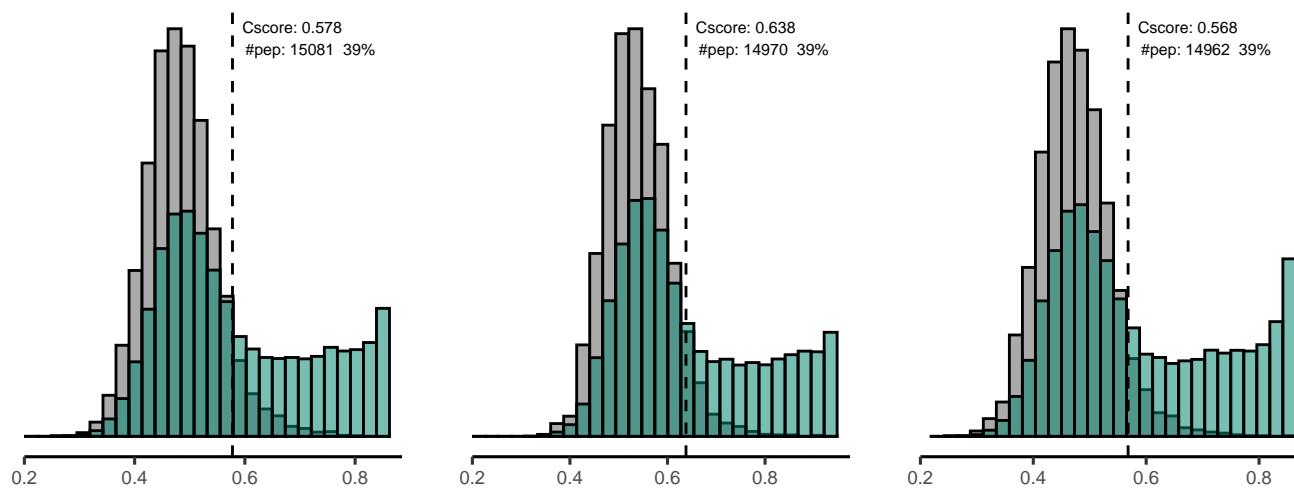
18\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD128180618\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD128180618\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD1xS1fM1\_



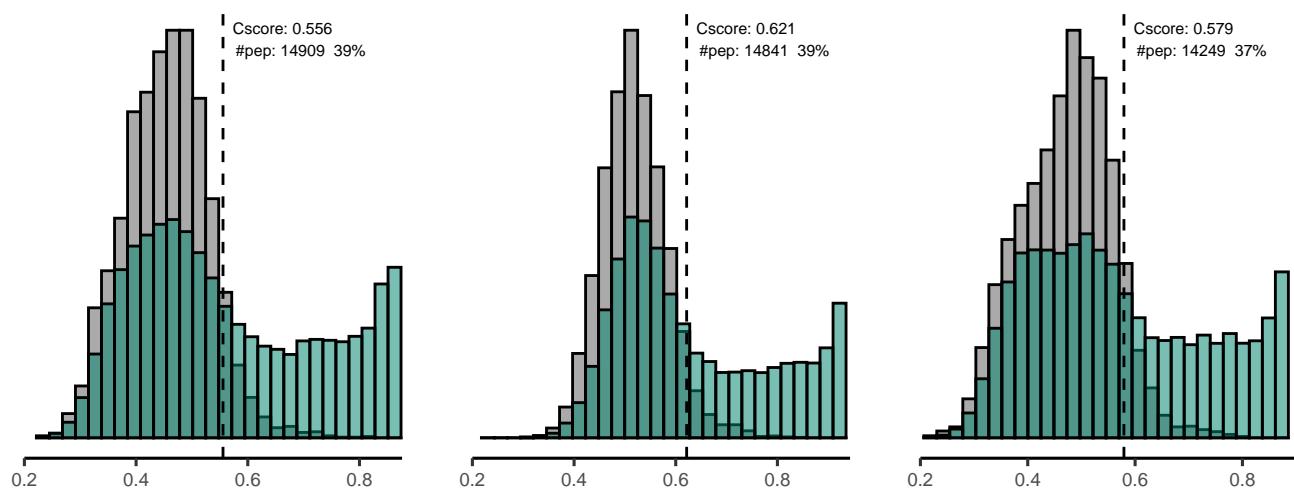
18\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD128180618\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD1xS1fM1\_20190228\_QX1\_PhGe\_SA\_EASY12-10\_Deigendesch\_G7\_



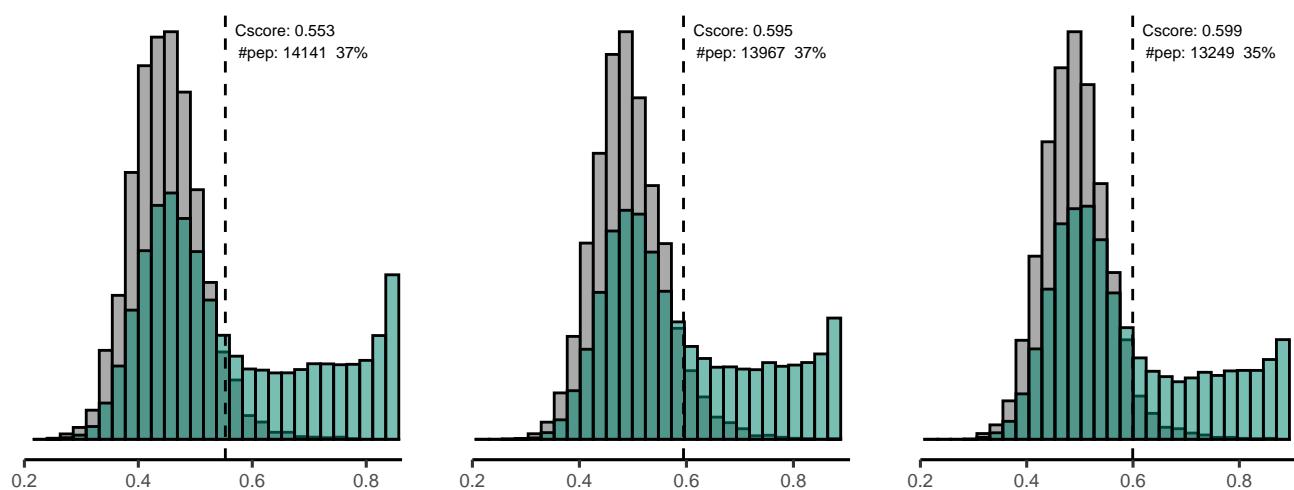
18\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD128180618\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD128180618\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD1xS1fM1.



228\_QX1\_PhGe\_SA\_EASY12-10\_Deigendes20180618\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_82018028010X1\_PhGe\_SA\_EASY12-10\_Deigendesch\_G9\_828799.



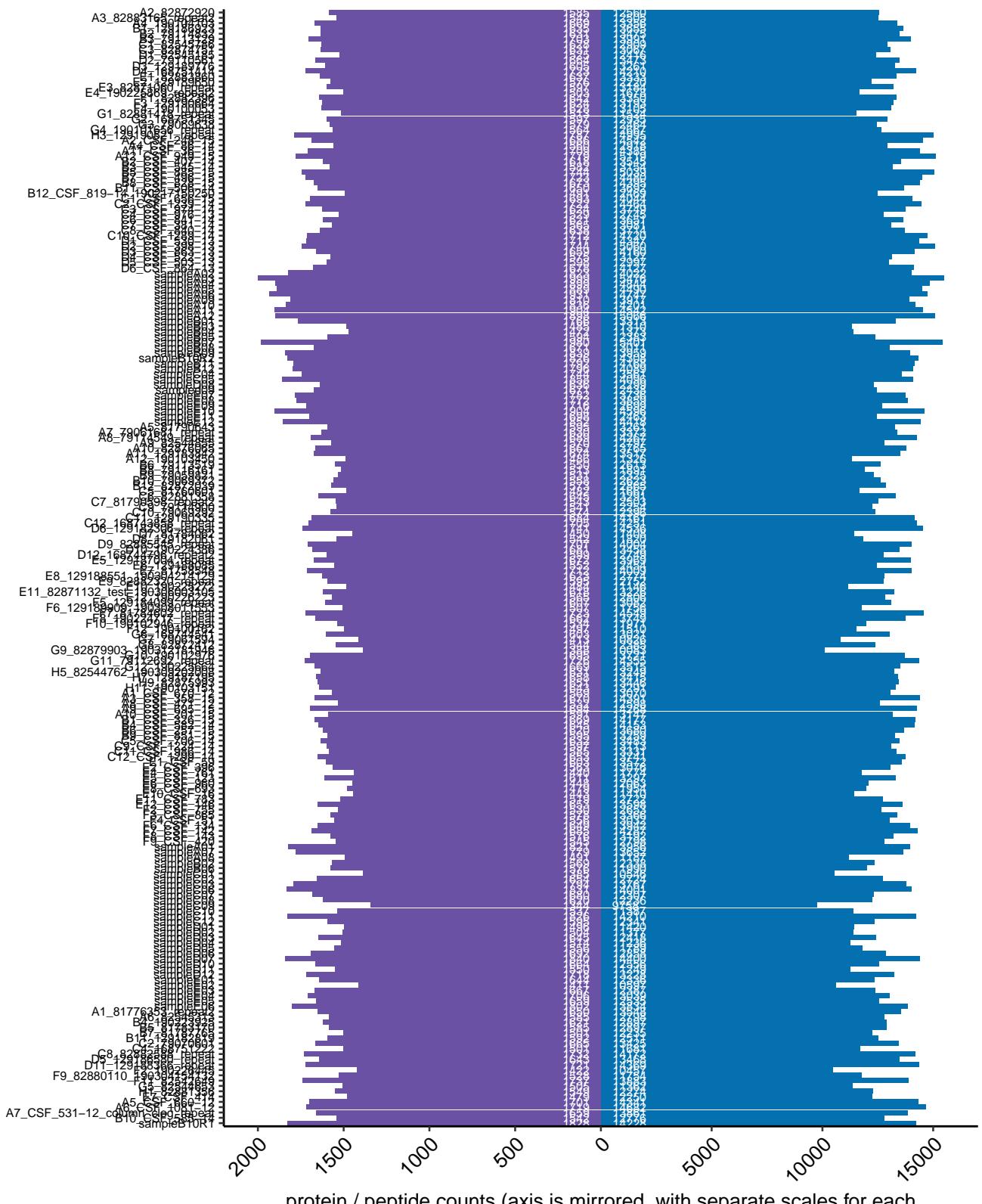
18\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD128180618\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD128180618\_QX0\_JaBa\_SA\_LC12\_5\_CSF1\_1\_8-1xD1xS1fM1.



## 1.2 number of peptides and proteins

These plots show the number of (target) peptides that are ‘detected’ per sample. For DDA, ‘detected’ implies the peptide has a MS/MS identification. Peptides quantified through match-between-runs (MBR) are quantified but not detected/identified. In case of DDA, we also show the number of peptides quantified through MBR. For DIA, we refer to a peptide as ‘detected’ if the confidence score (for identification) is  $\leq 0.01$ .

Samples in this plot are sorted by their experimental group, and then ordered and by their name within each group. This data is also available in the output table ‘samples.xlsx’.



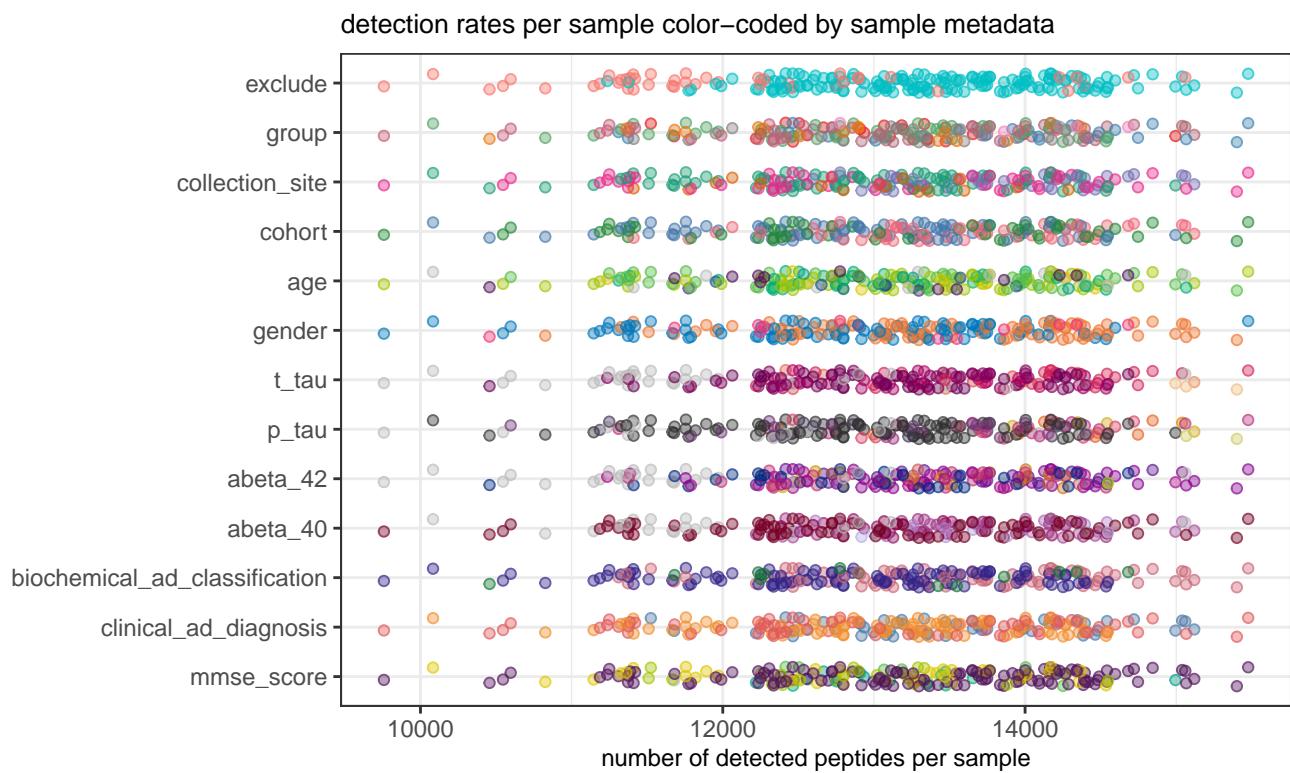
█ peptides: identified & quantified  
█ proteins: identified & quantified

### 1.2.1 color-coding sample metadata

The number of detected peptides in a sample, as compared to other samples within a dataset, can be used as a measure for sample quality. Color-coding individual samples for metadata that you provided as input (e.g. experiment batch, sample handling order, gel lanes, etc.) allows visual inspection as to whether these relate to the rate of successful peptide detection.

The figure below provides an overview of all sample metadata at a first glance. On each row all samples in the dataset are shown as a data point, each color-coded by the respective property shown on the y-axis (with minor vertical jitter for visual clarity). If any of these metadata coincide with a major effect on the number of detected peptides, this should become apparent by a clustering of samples by color-code. Hereafter, an additional set of figures will further expand this overview into detailed figures for each sample property.

Note that the visualization of sample metadata in this report depends on user-provided input; each column in the metadata input table (besides sample names) that contains more than 1 unique value is automatically used as a factor for color-coding all figures in this section. All information shown in these figures is also available in the output table ‘samples.xlsx’.



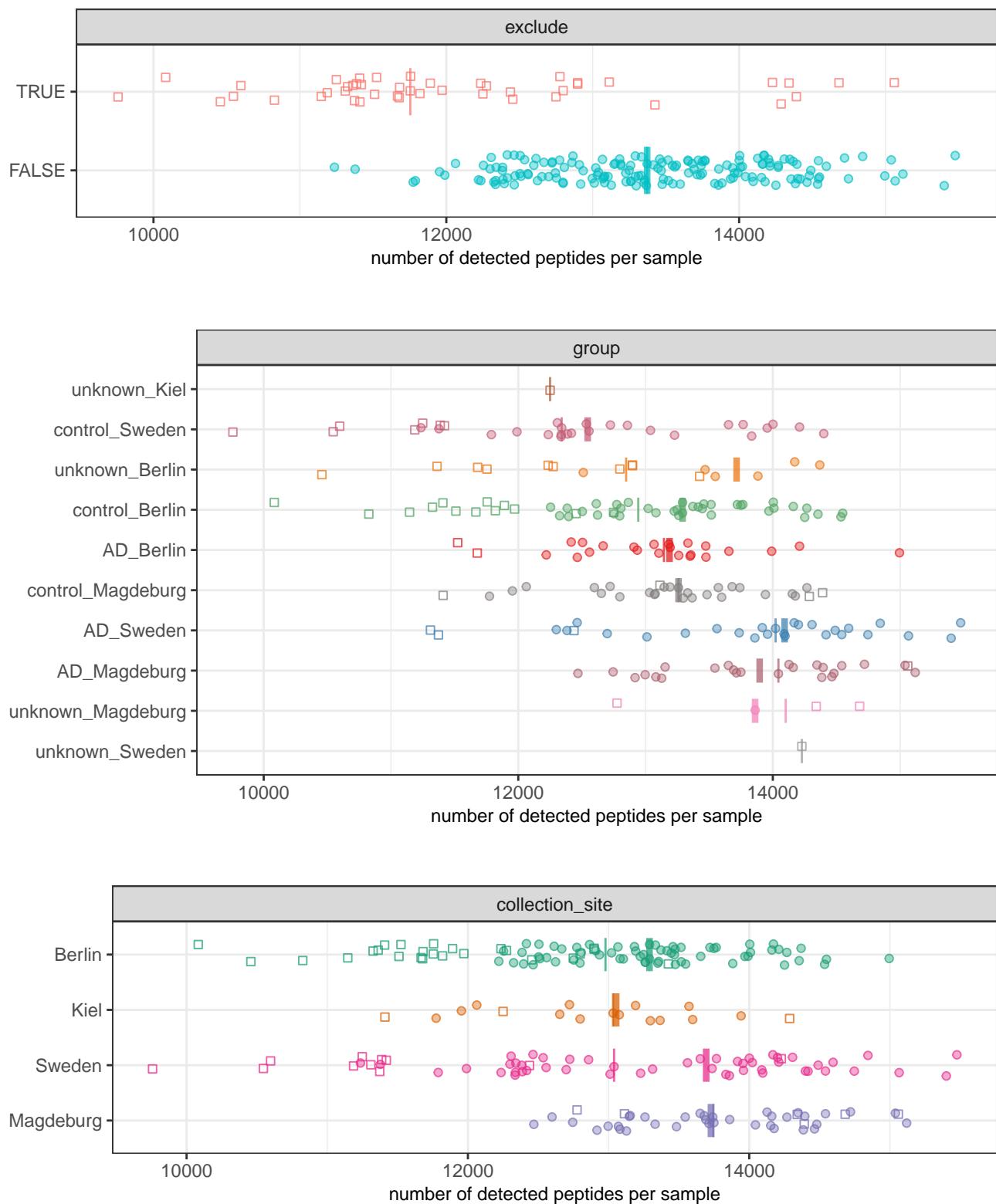
### color-coding sample metadata, expanded

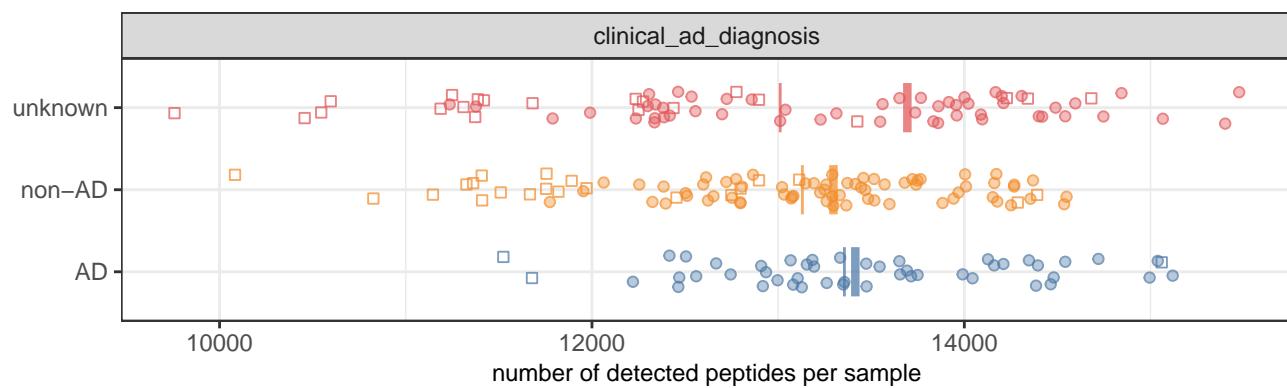
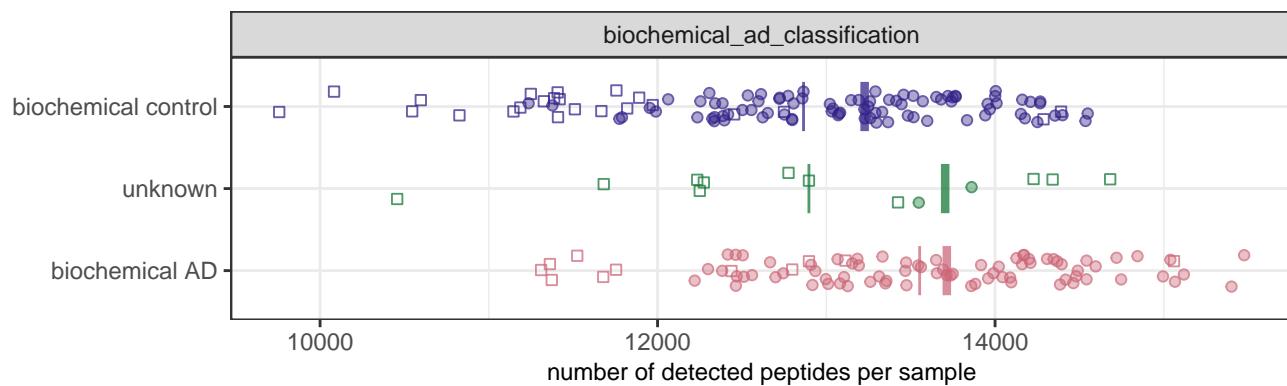
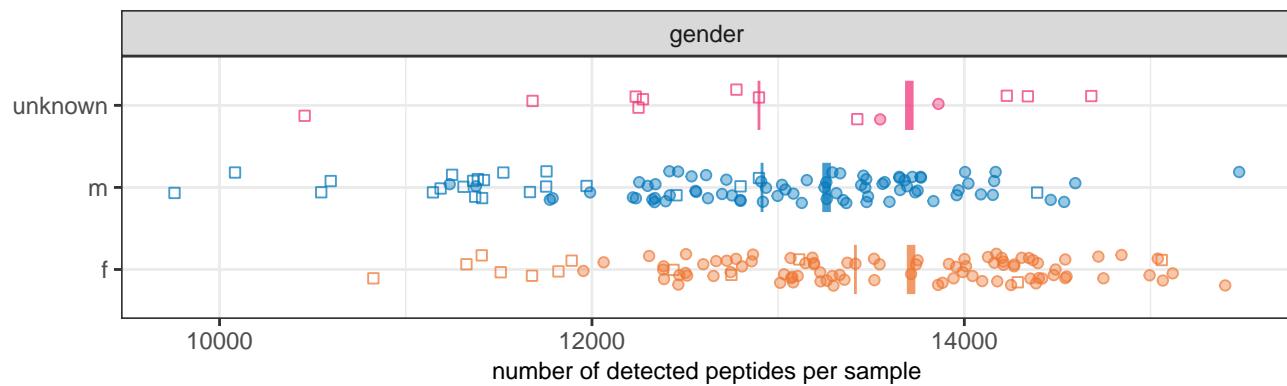
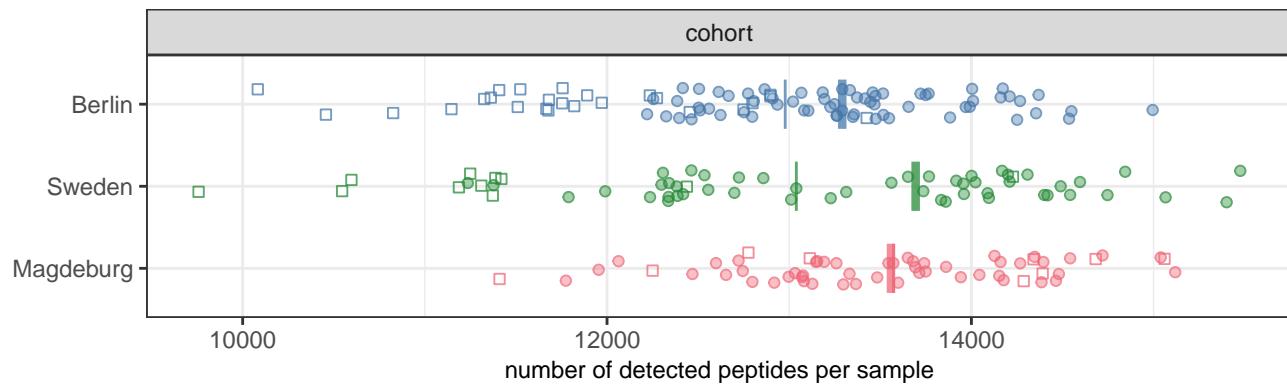
To further detail each sample property, each row in the above figure is now split into separate plots. Thus, a figure is generated for each property in the user-provided metadata (column in the samples table, its name shown in the plot title).

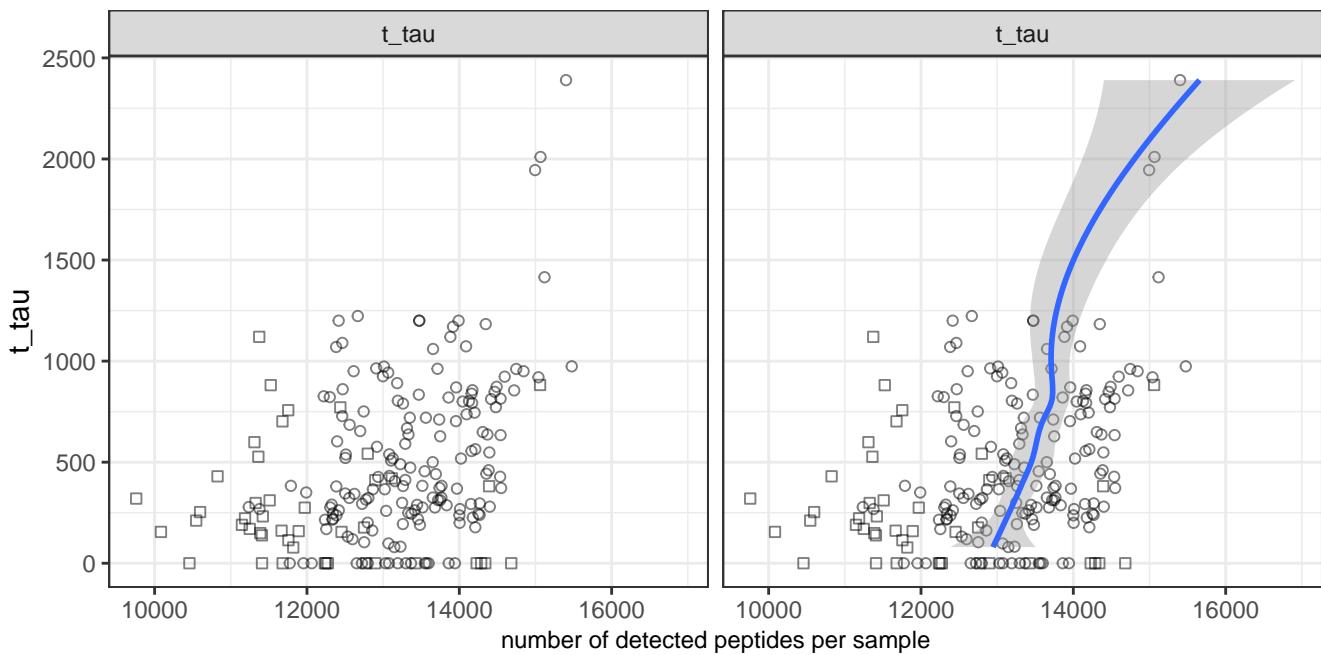
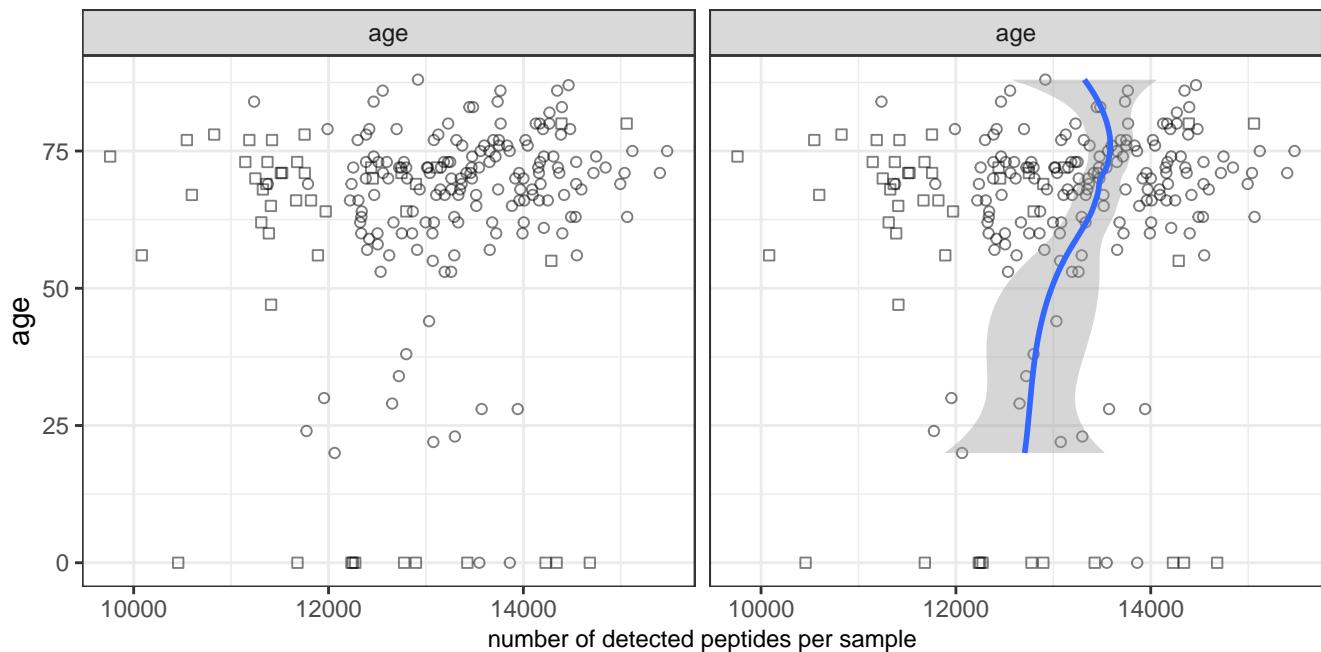
For categorical variables, a scatterplot shows on the y-axis all unique variables while the x-axis depicts the number of detected peptides. Colors are consistent with the above plot. *exclude* samples, if any, are depicted as squares. The median value is shown as a vertical line (thin line = median over all samples, wider line = median while discarding *exclude* samples). For continuous variables, a scatterplot without (left panel) and with Loess fit is shown (right panel, visualized as blue line if data was successfully fitted).

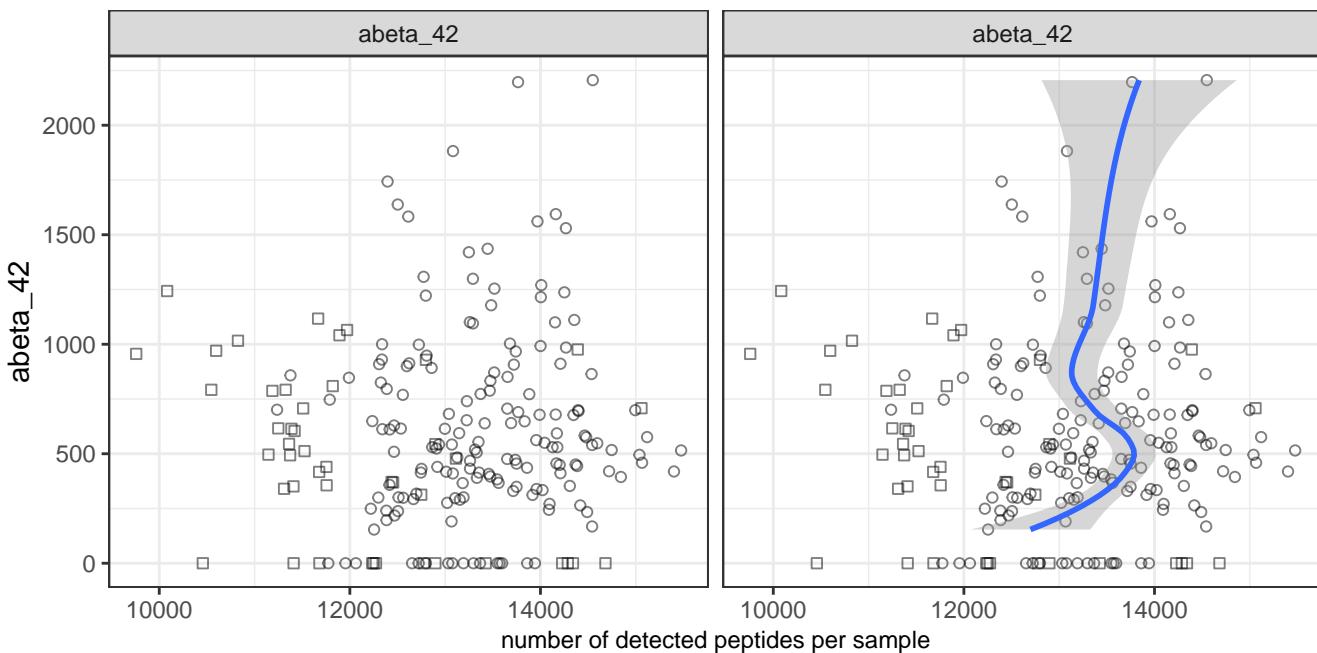
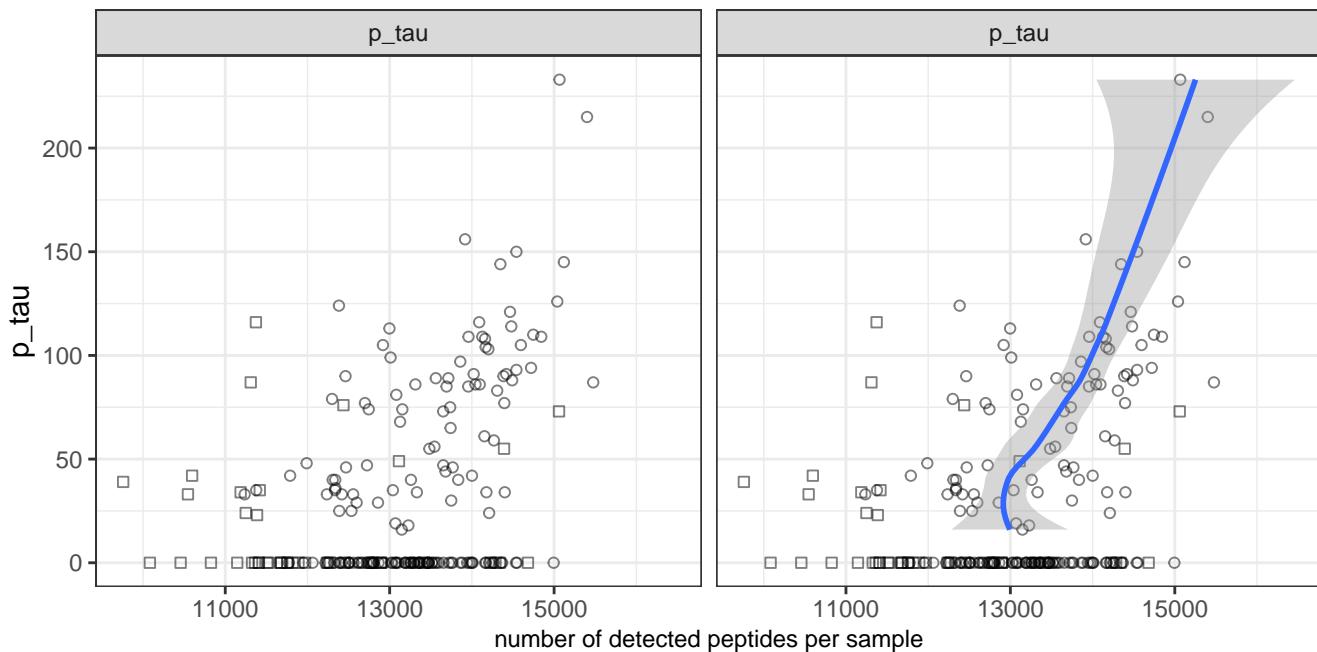
Note that samples flagged as *exclude* are user-provided in the sample metadata table. These are included in data visualizations but excluded from downstream statistical analysis (later part of the report).

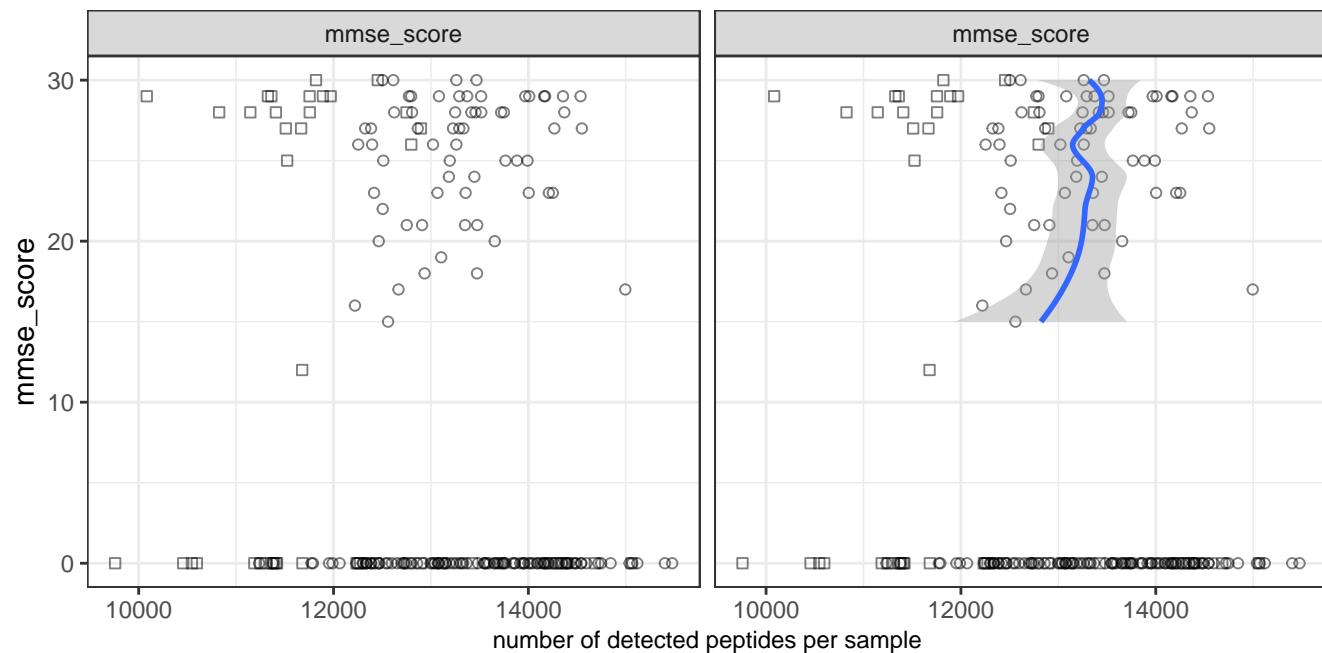
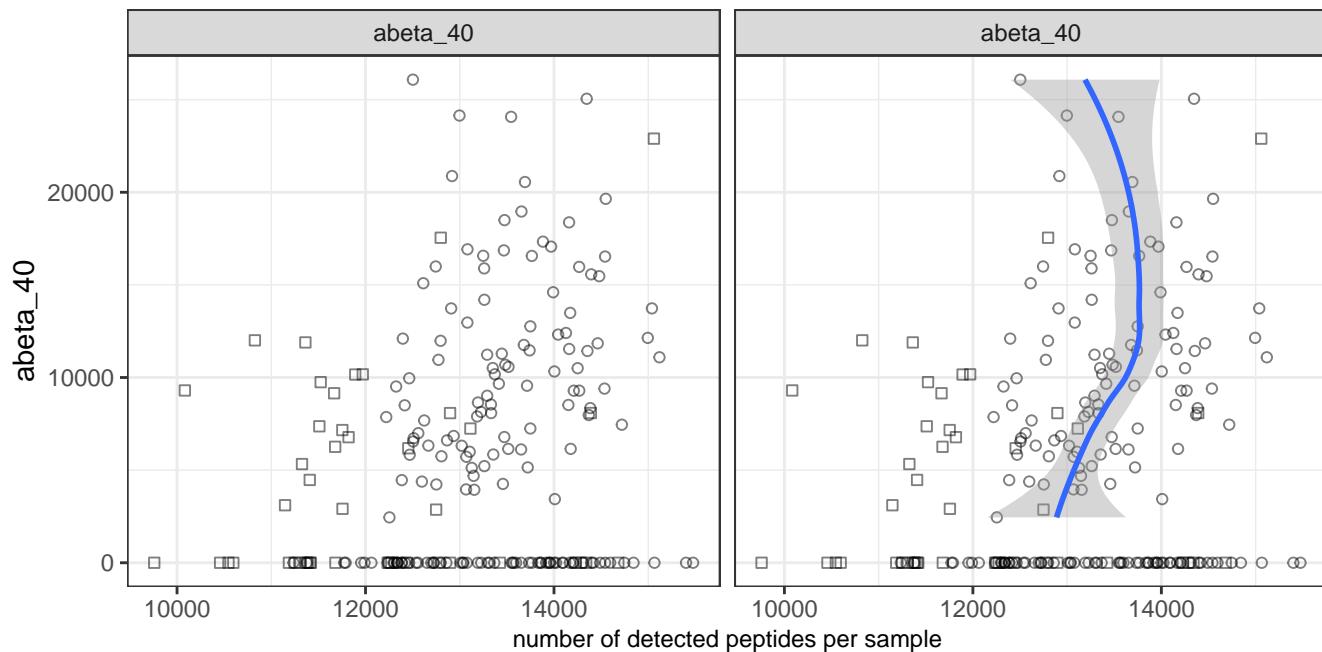
For example: the first plot shows color-coding by the ‘group’ property, so each row represents a sample group. If samples in a particular group systematically yield fewer peptides than another group, a clear pattern will be visible.







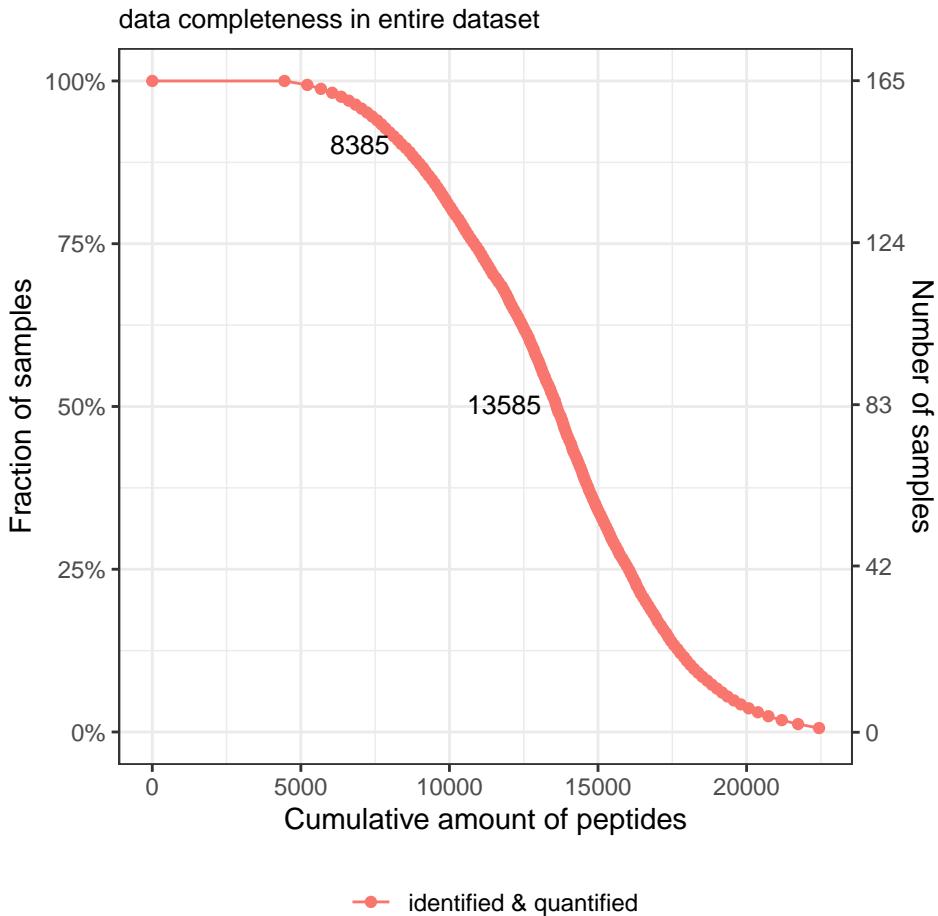




## 1.3 data completeness

To visualize how many peptides are consistently identified in multiple samples, the first figure summarizes how common missing values are in the entire dataset. Optimally, most peptides are identified in 100% of samples and this curve slowly falls off. The following figure shows for each sample whether its peptides are also present in other samples in the dataset or whether these are unique to a (minor) subset of samples. You can use this mark of experimental consistency to compare datasets generated by similar protocols and mass-spec acquisition.

### 1.3.1 cumulative distribution



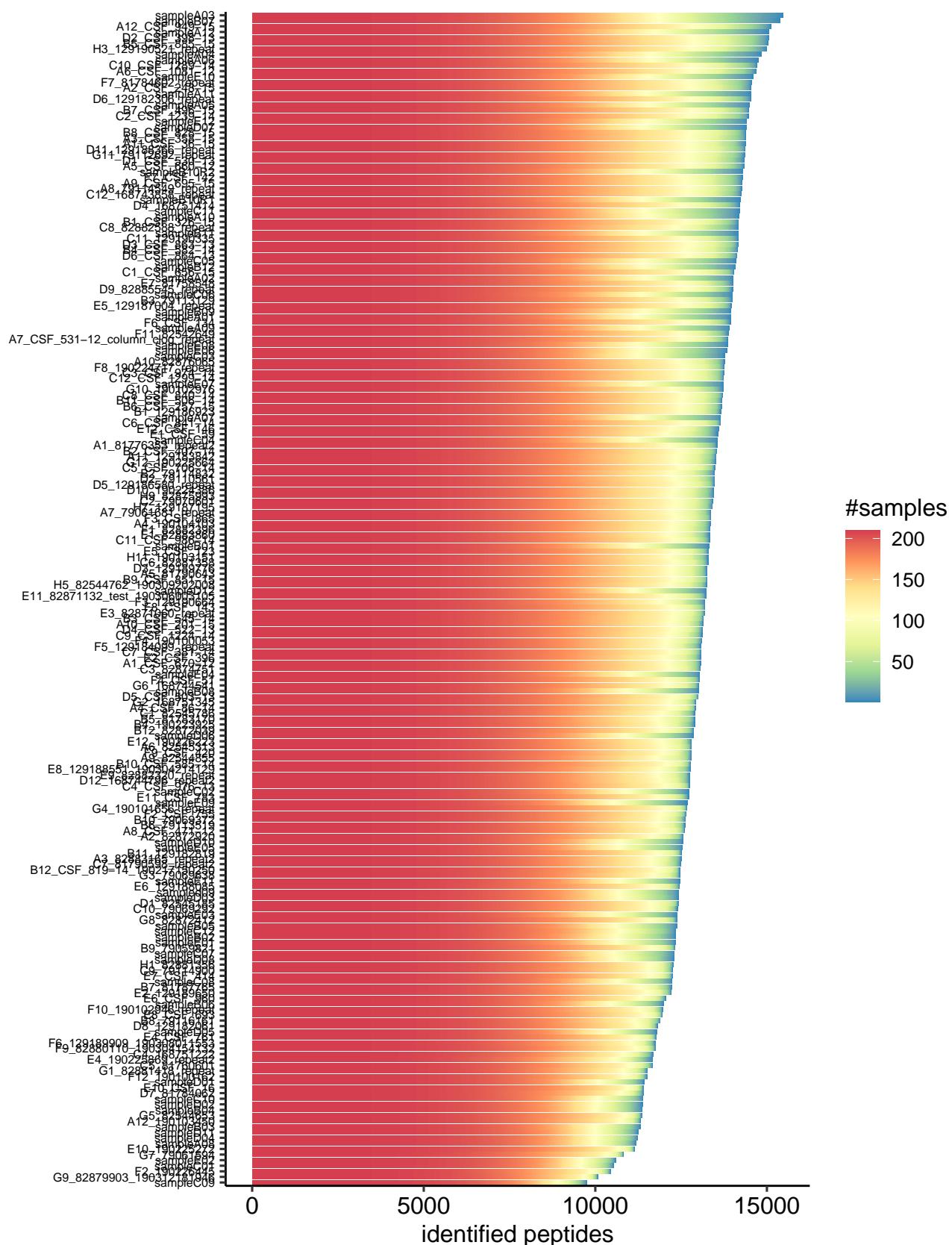
Samples flagged as ‘exclude’ (by user) are not taken into account in this figure. Exact values are shown for data points matching 90% and 50% of samples to convenience comparison between analyses (e.g. before/after configuring ‘exclude’ samples, or comparing between experiments of similar protocol).

### 1.3.2 peptide detection frequency

Each identified peptide in a sample is classified and color-coded by the number of other samples where the same peptide is present. Visualization of the amount of peptides that overlap with other samples in the dataset, from peptides identified in most samples (red) to one-hit-wonders (blue), helps identify uncommon samples (more blue/green than other samples).

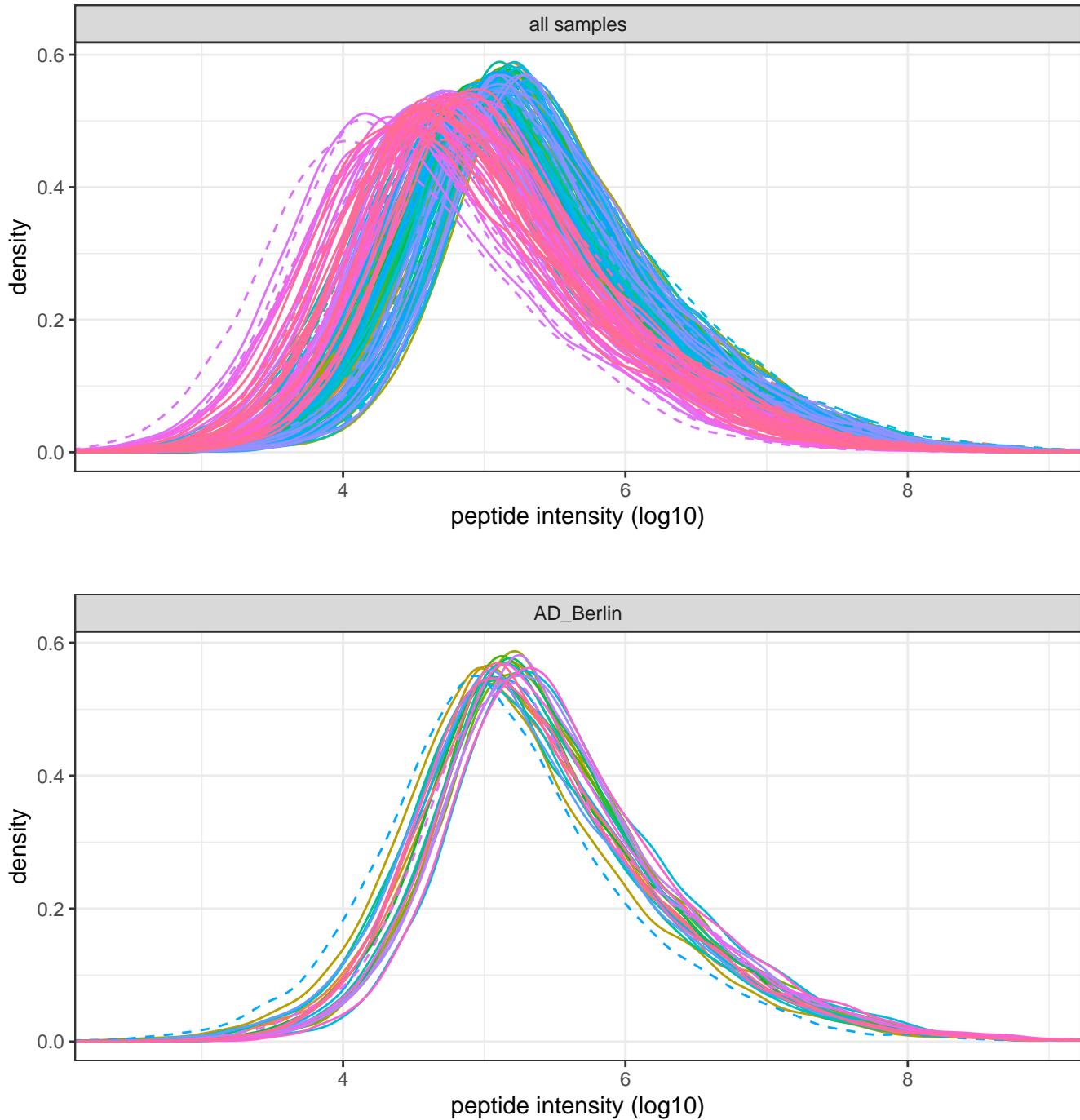
Optimally, the majority of peptides in each sample are red~orange with relatively few uniquely identified peptides (blue~green). Samples are sorted by the total amount of detected peptides.

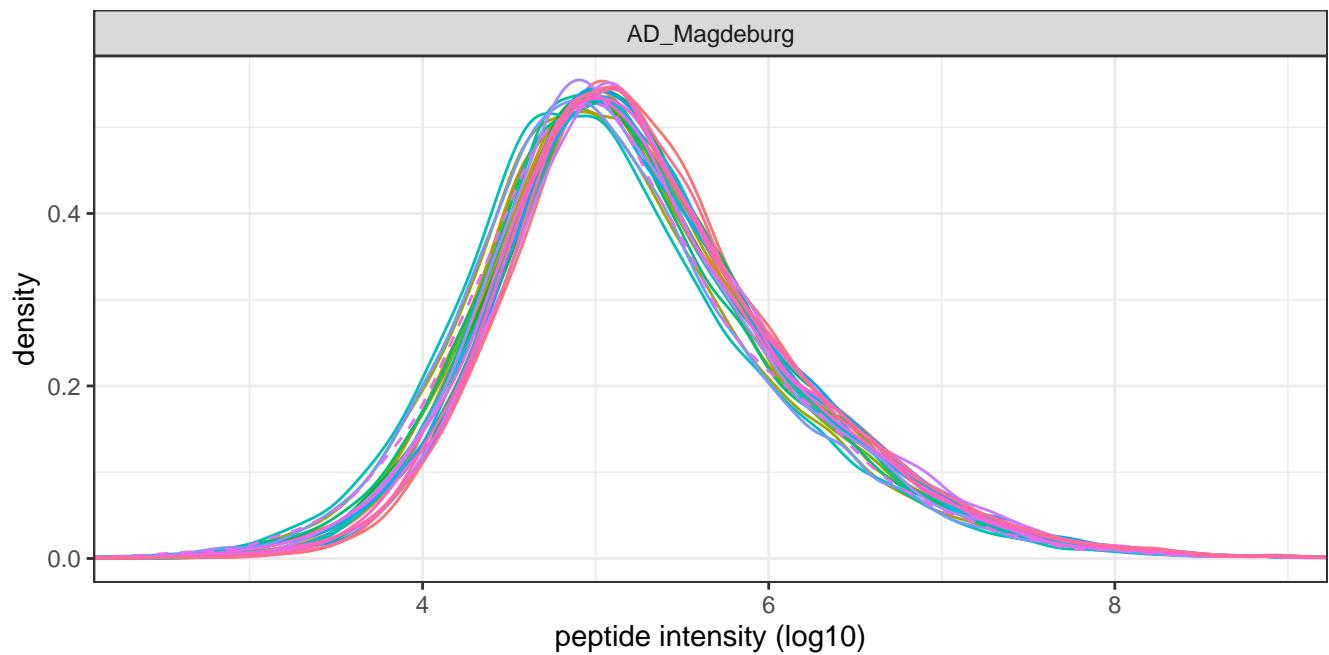
Number of samples in which a peptide is identified vs presence in individual sample



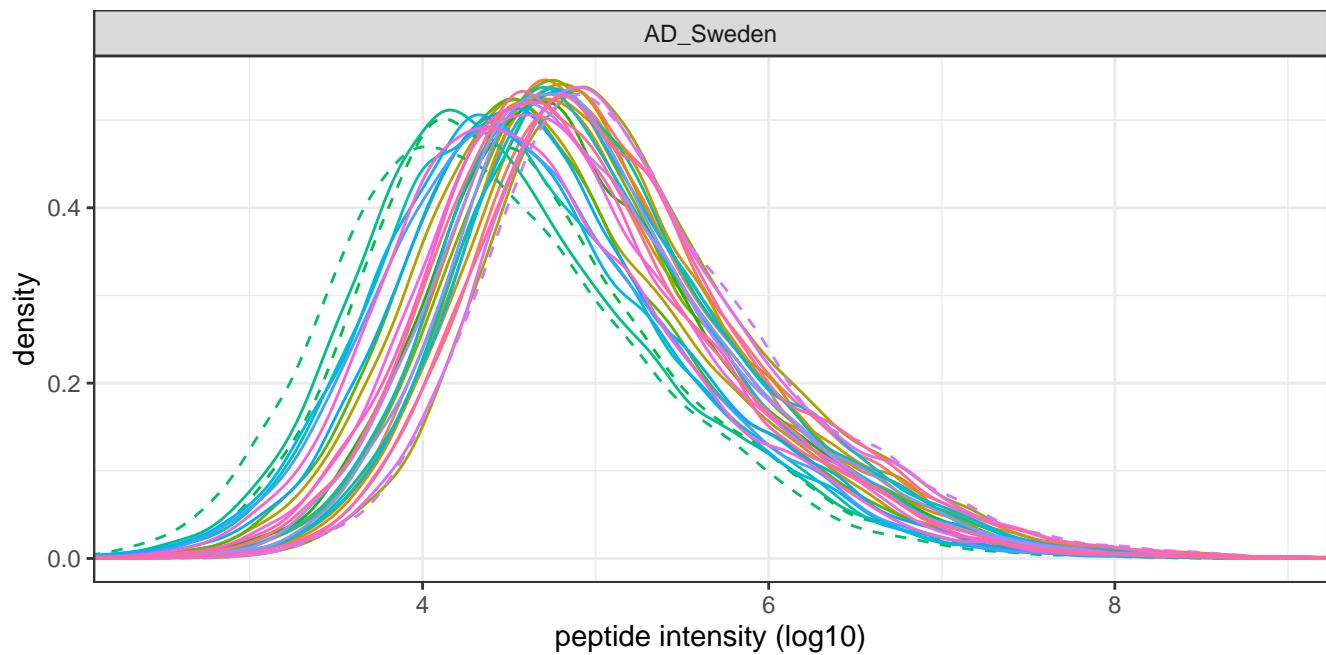
## 1.4 abundance distributions

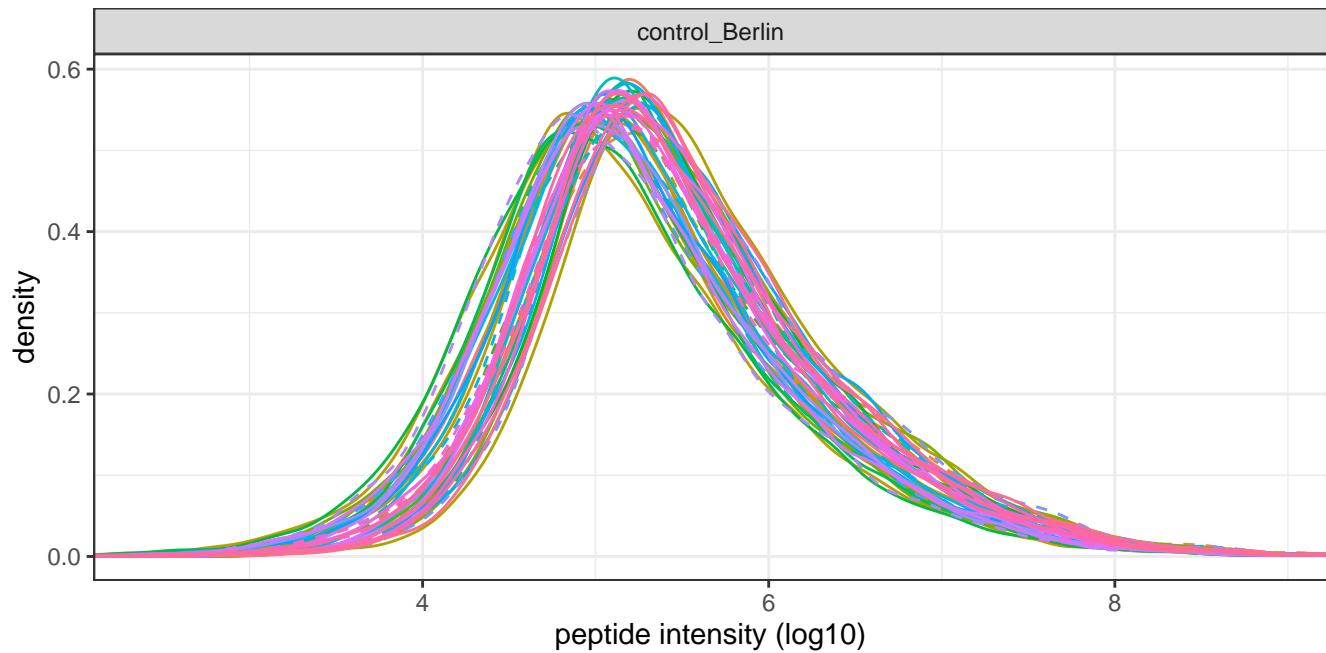
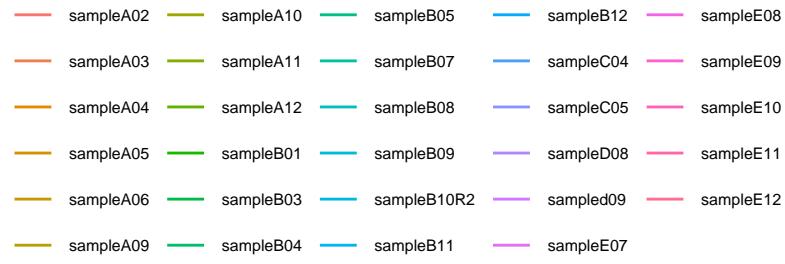
The figures in this subsection are used to identify unexpected mass-spec sensitivity or sample loading differences. Peptide data is shown as provided in input files, so peptide filtering nor intensity normalization has been applied yet (for proper QC, make sure the software that generated the input data did not apply normalization prior). If the dataset is DDA, match-between-runs (MBR) peptides are included in these distributions whereas for DIA only ‘detected’ peptides (based on confidence score threshold) are included.



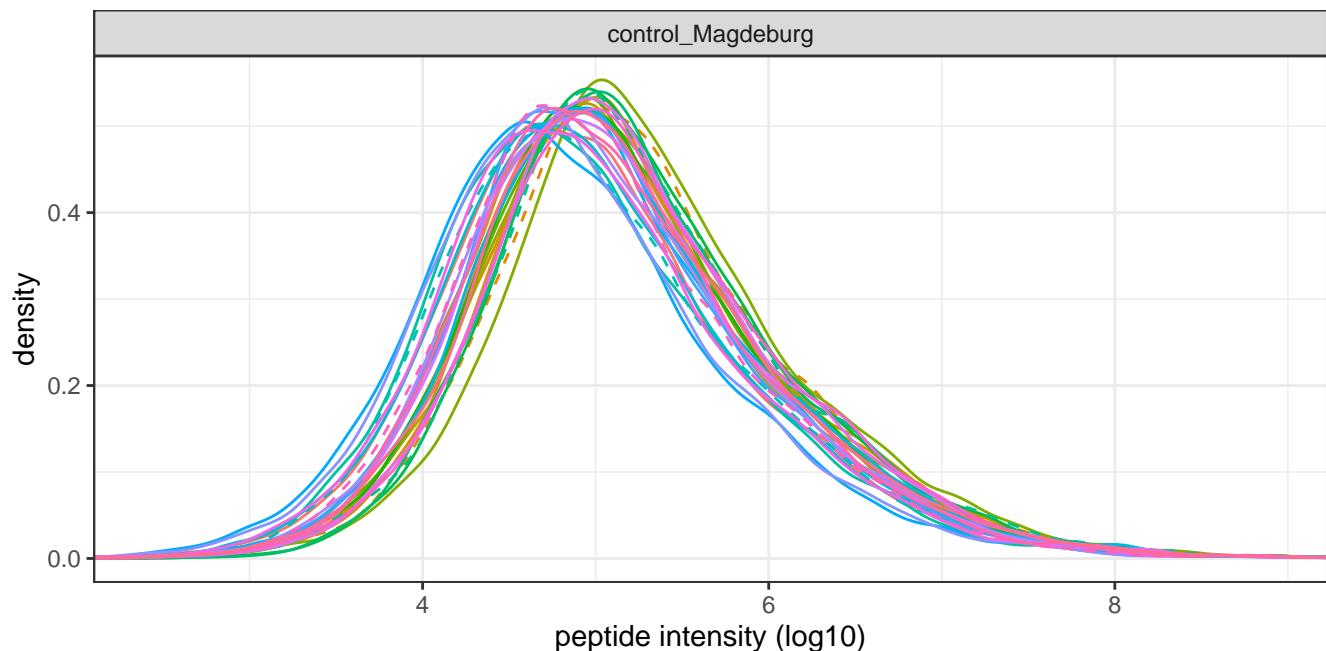


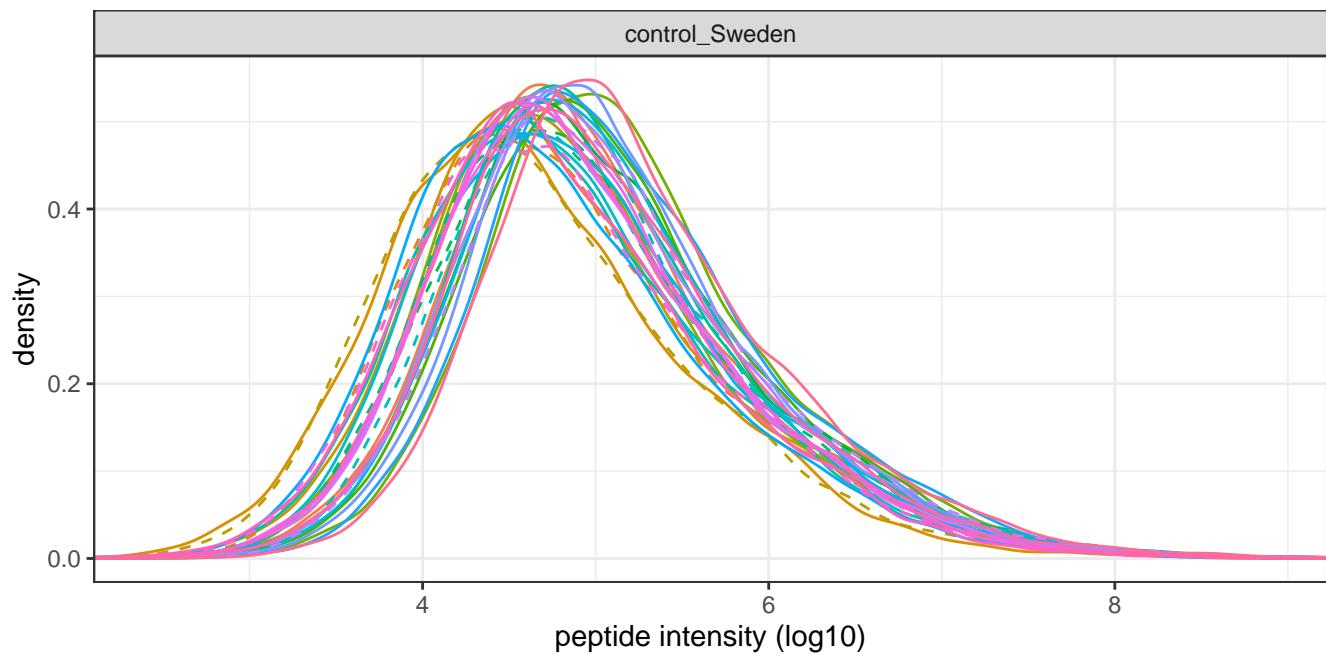
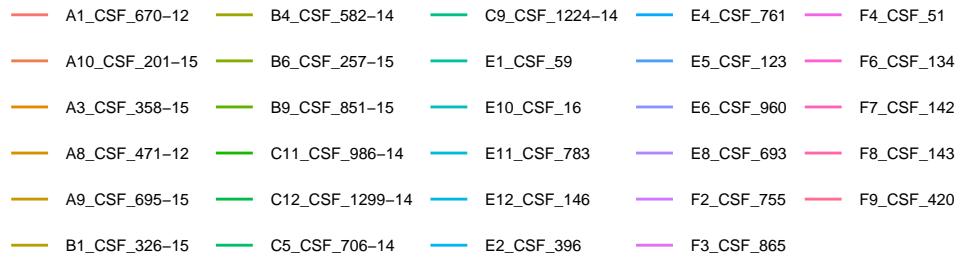
A11_CSF_36-15	B12_CSF_819-14_190217150250	B8_CSF_828-15	C4_CSF_976-13	D2_CSF_398-13
A12_CSF_949-15	B2_CSF_407-14	C1_CSF_656-15	C6_CSF_841-14	D3_CSF_863-13
A2_CSF_248-15	B3_CSF_545-14	C10_CSF_1289-14	C7_CSF_381-14	D4_CSF_522-13
A4_CSF_86-14	B5_CSF_885-15	C2_CSF_1239-14	C8_CSF_840-14	D5_CSF_503-13
B11_CSF_506-14	B7_CSF_496-15	C3_CSF_974-14	D1_CSF_530-13	D6_CSF_864-13

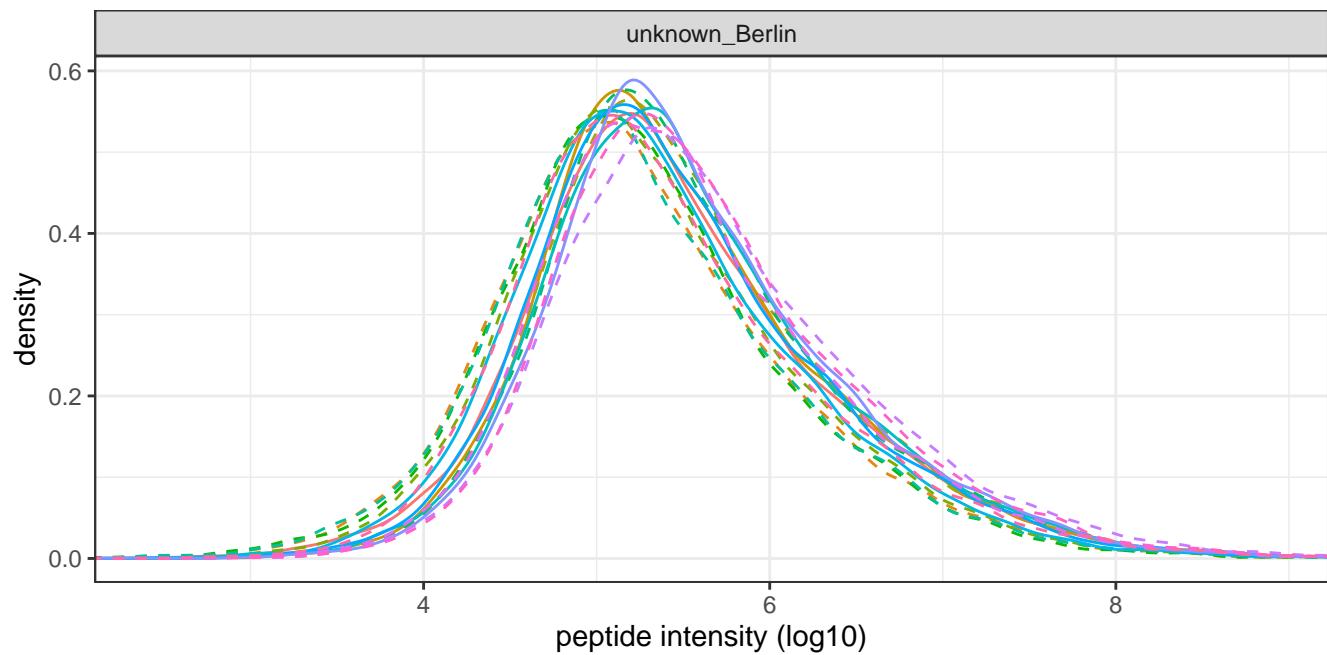




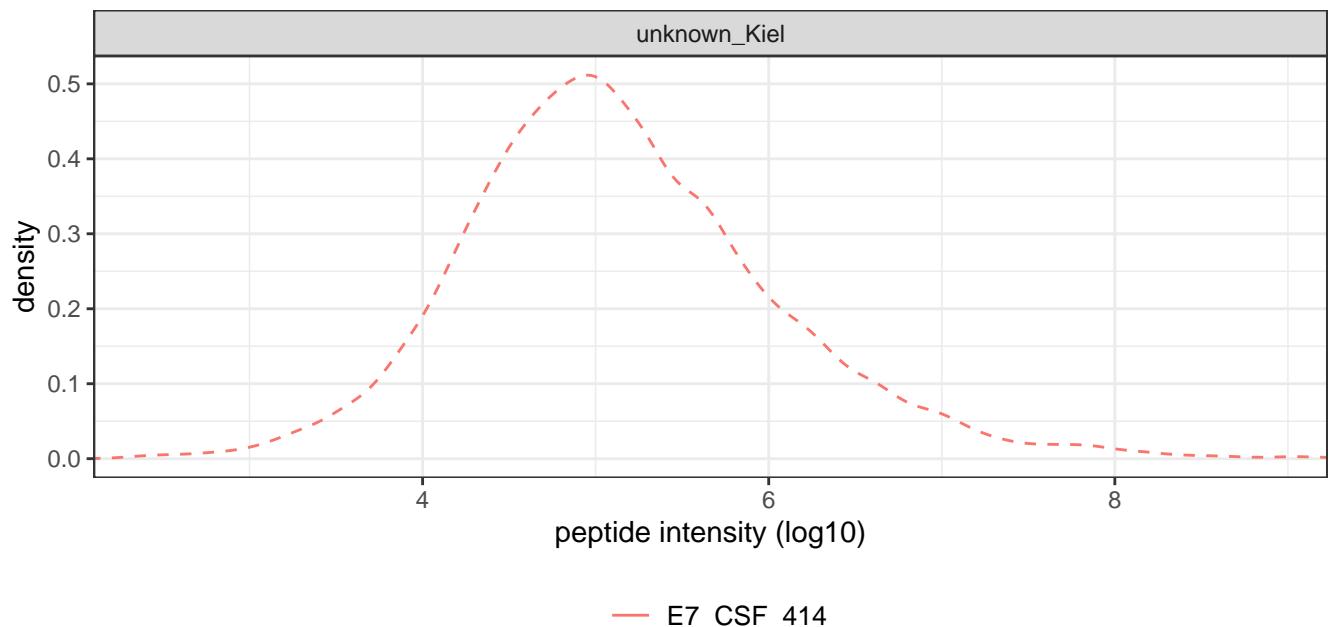
A10_82876065	B8_79116161	D12_168744796_repeat2	E7_81758548	G11_79112692_repeat
A11_129183842	B9_79059821	D6_129182306_repeat	E8_129188551_190304214129	G12_190225664
A12_190103450	C10_79069292	D7_81784062	E9_82882320_repeat	G6_168744541
A5_81790643	C11_129190335	D8_129182061	F10_190102946_repeat	G7_79061594
A7_79061681_repeat	C12_168743858_repeat	D9_82885545_repeat	F12_190100162	G8_82872412
A8_79114549_repeat	C5_81760601	E10_190225272	F5_129184099_repeat	G9_82879903_1903121815
A9_82544855	C6_82881358	E11_82871132_test_190306003105	F6_129189909_190308011553	H11_190103151
B10_79069372	C7_81790598_repeat2	E12_190226223	F7_81784602_repeat	H5_82544762_190309202C
B12_82872039	C9_79114900	E5_129187004_repeat	F8_190224717_repeat	H7_129187195
B6_79113519	D10_190224386	E6_129188085	G10_190102976	H9_82875993

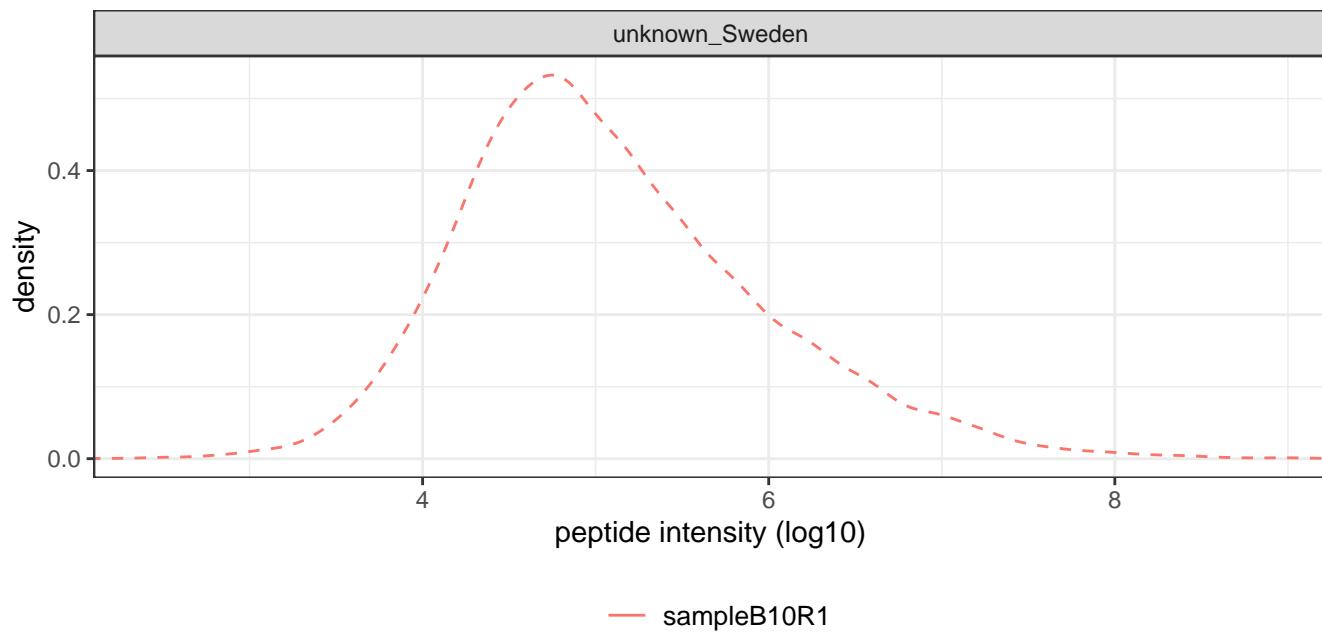
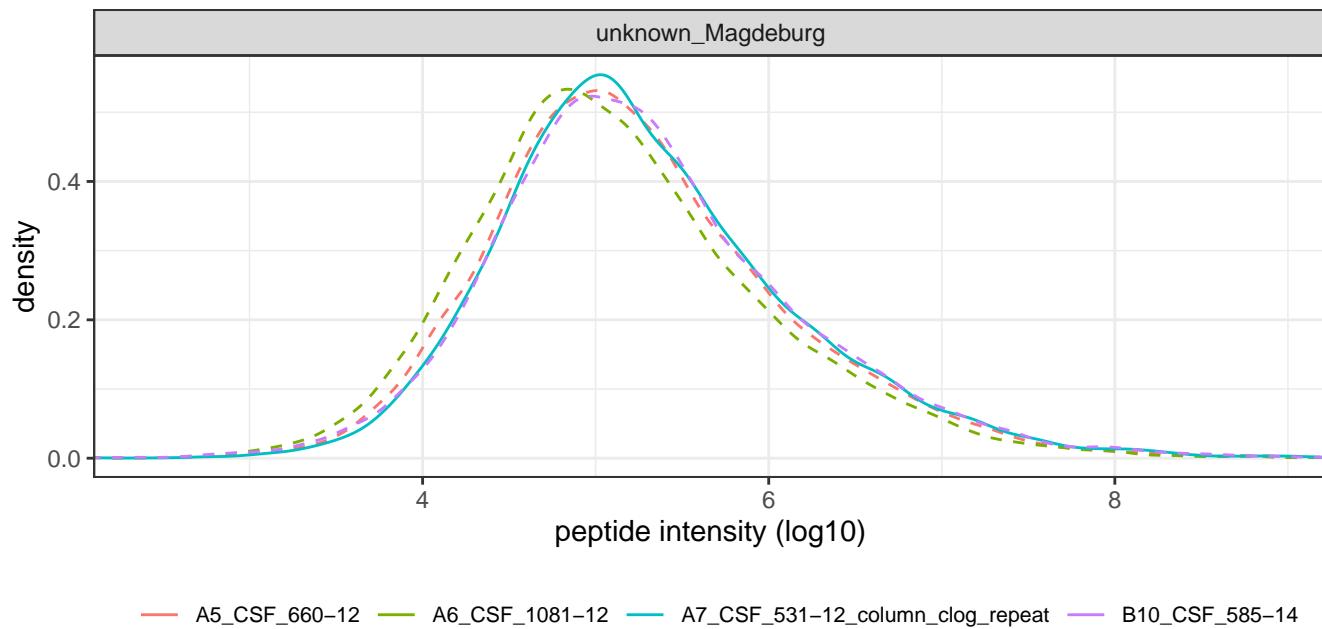






— A1\_81776353\_repeat2 — B5\_81783170 — C8\_82882588\_repeat — F2\_190226445  
— A6\_82545313 — B7\_81787765 — D11\_129188366\_repeat — F9\_82880110\_190304154132  
— B11\_129182819 — C2\_79070601 — D5\_129186580\_repeat — G5\_82544653  
— B4\_190223925 — C4\_168751222 — F11\_82542649 — H1\_82881356



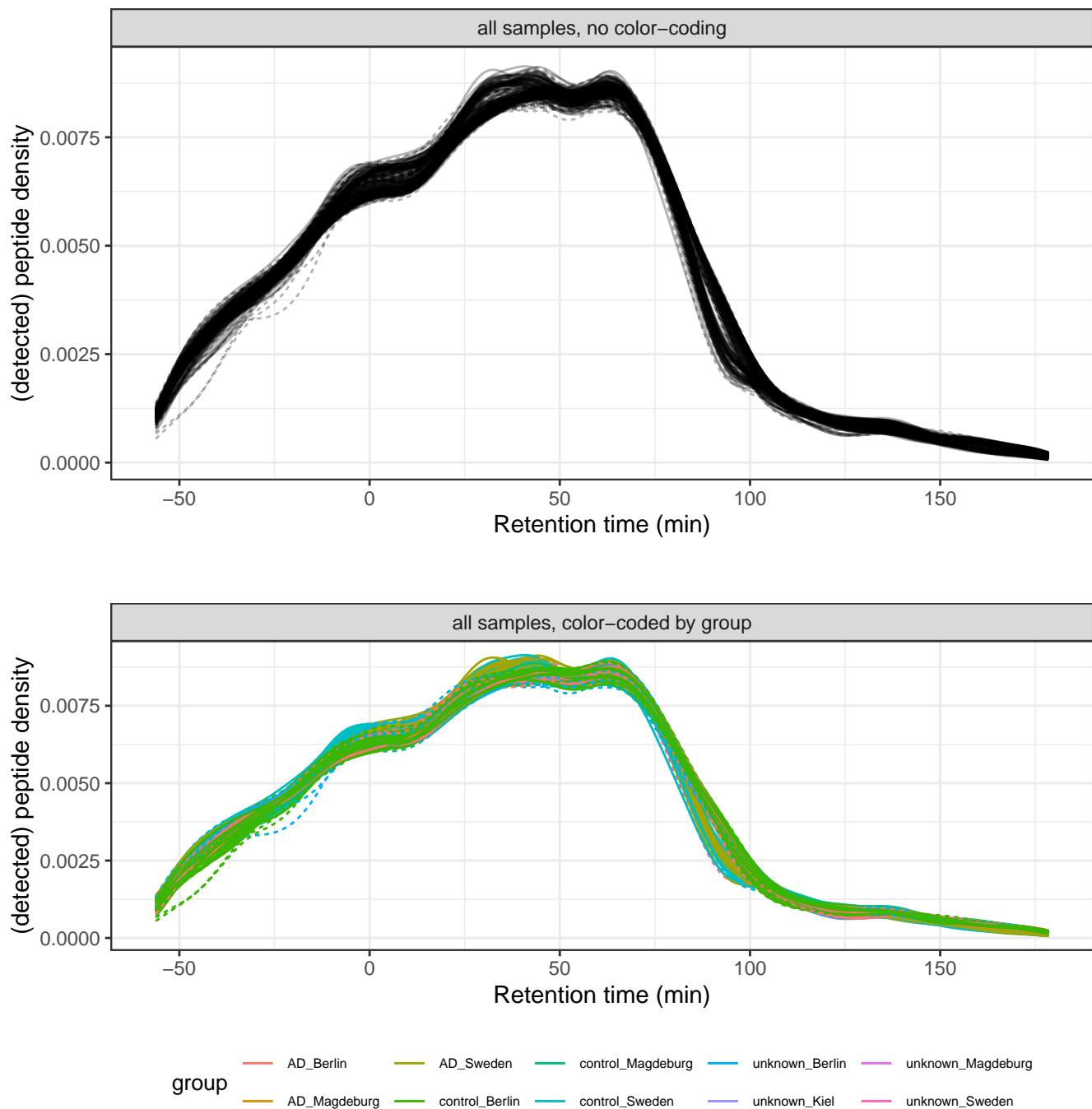


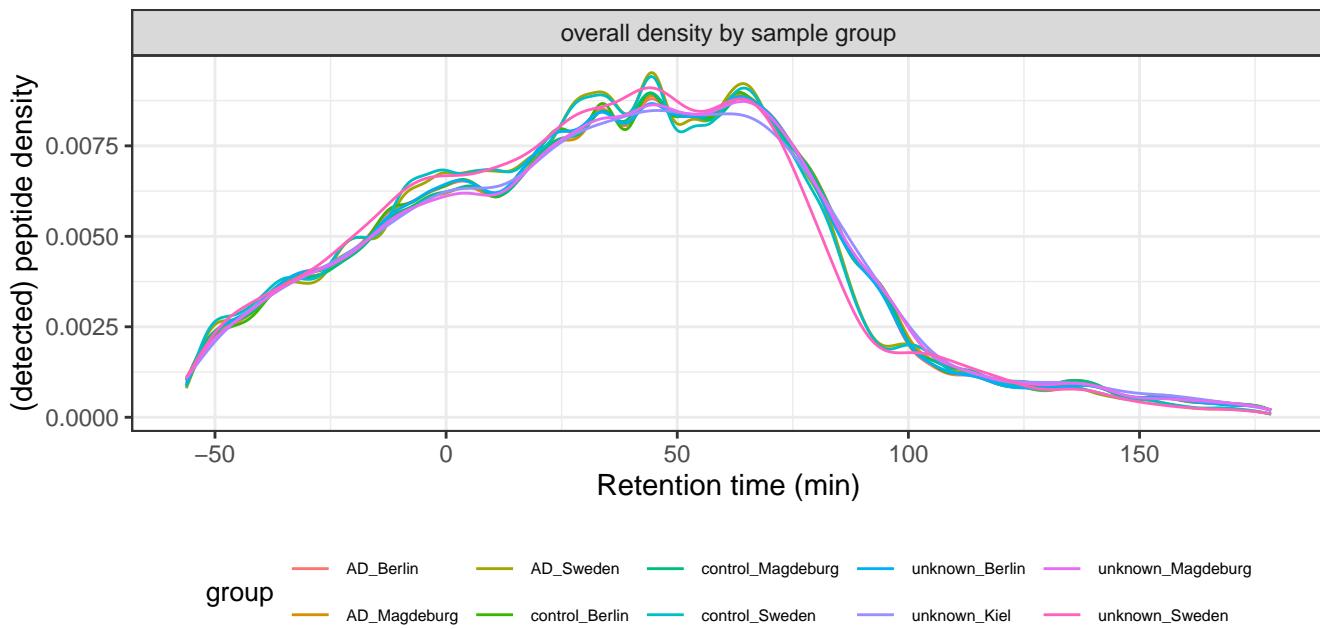
## 1.5 retention time

The figures in this section allow you to identify potential problems during HPLC elution, such as a temporarily blocking column, failing ionization spray or decreasing sensitivity over time. For each sample, all peptides that are also observed in a replicate (such that there is a point of reference available) are visualized.

### 1.5.1 retention time distributions

The density of the number of peptides eluting at each point in time. The figure below presents an overview of all samples that allows for the identification of outlier samples that follow distinct elution patterns. The following section shows details for each sample. Samples marked as ‘exclude’ in the provided sample metadata table are visualized as dashed lines.



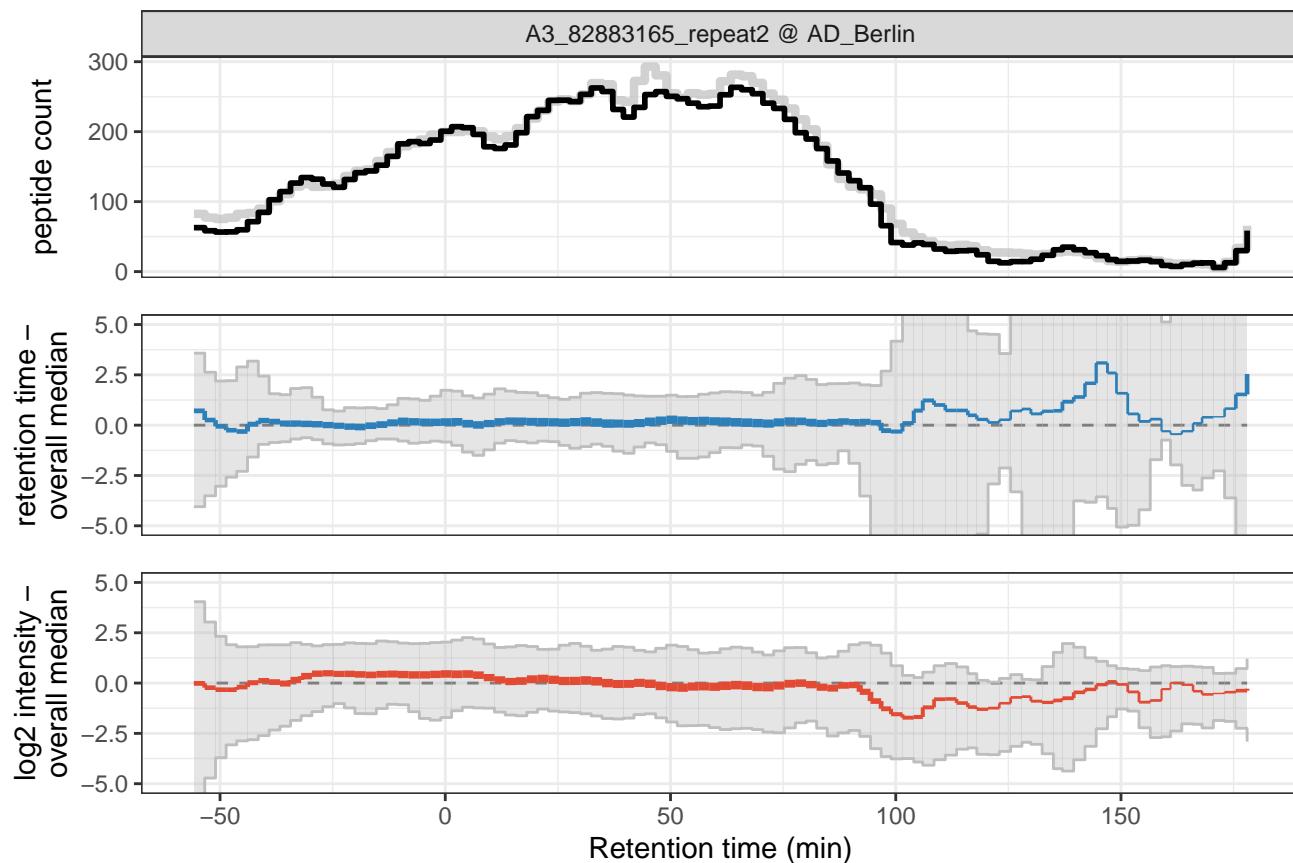
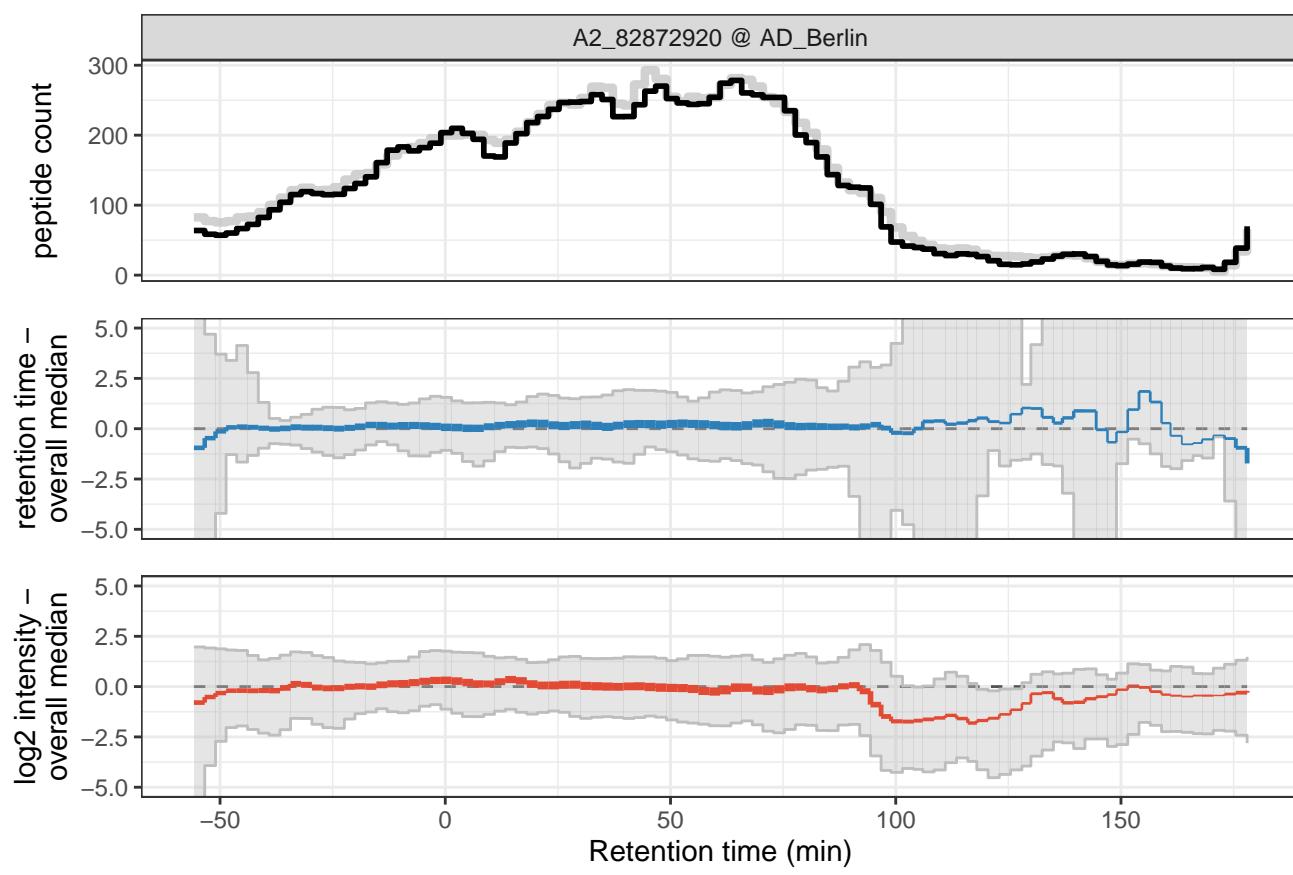


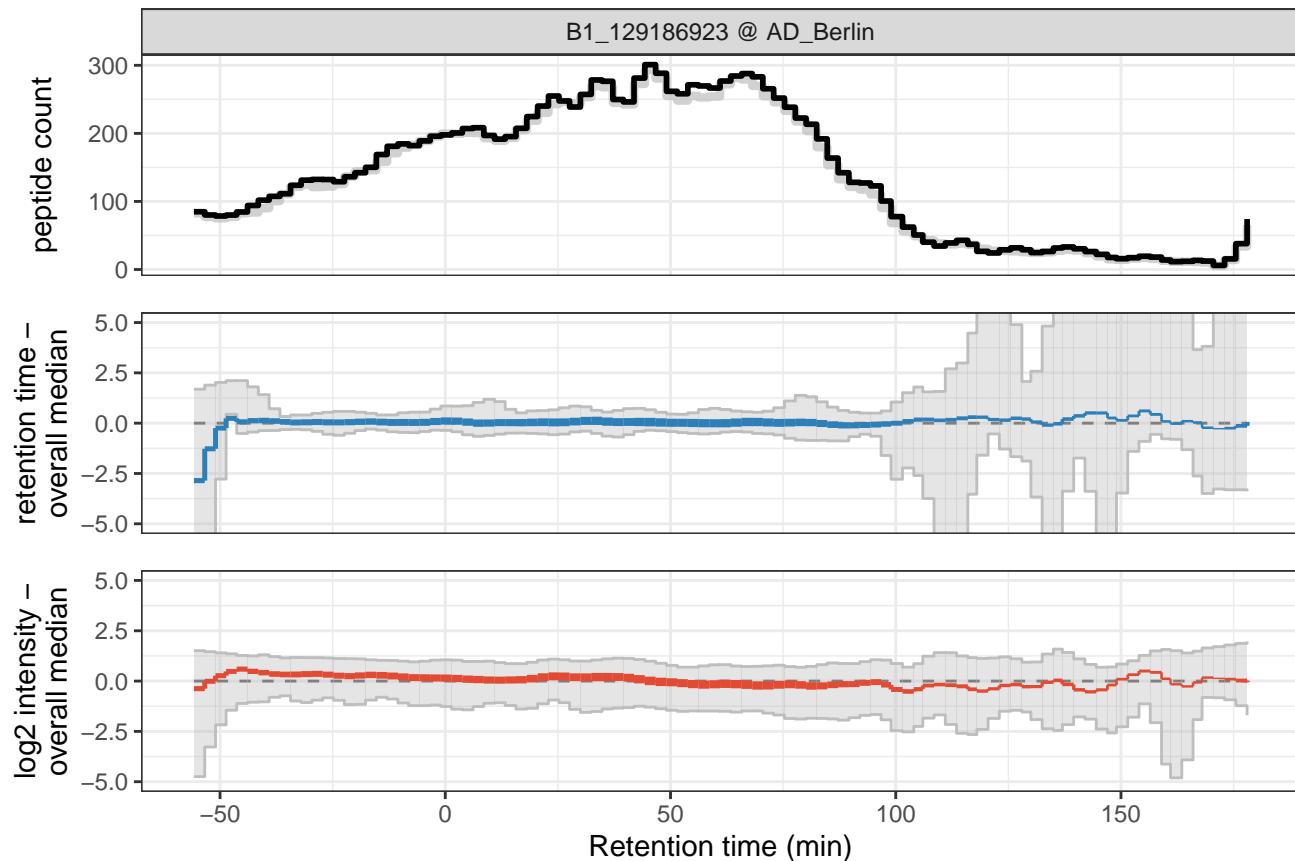
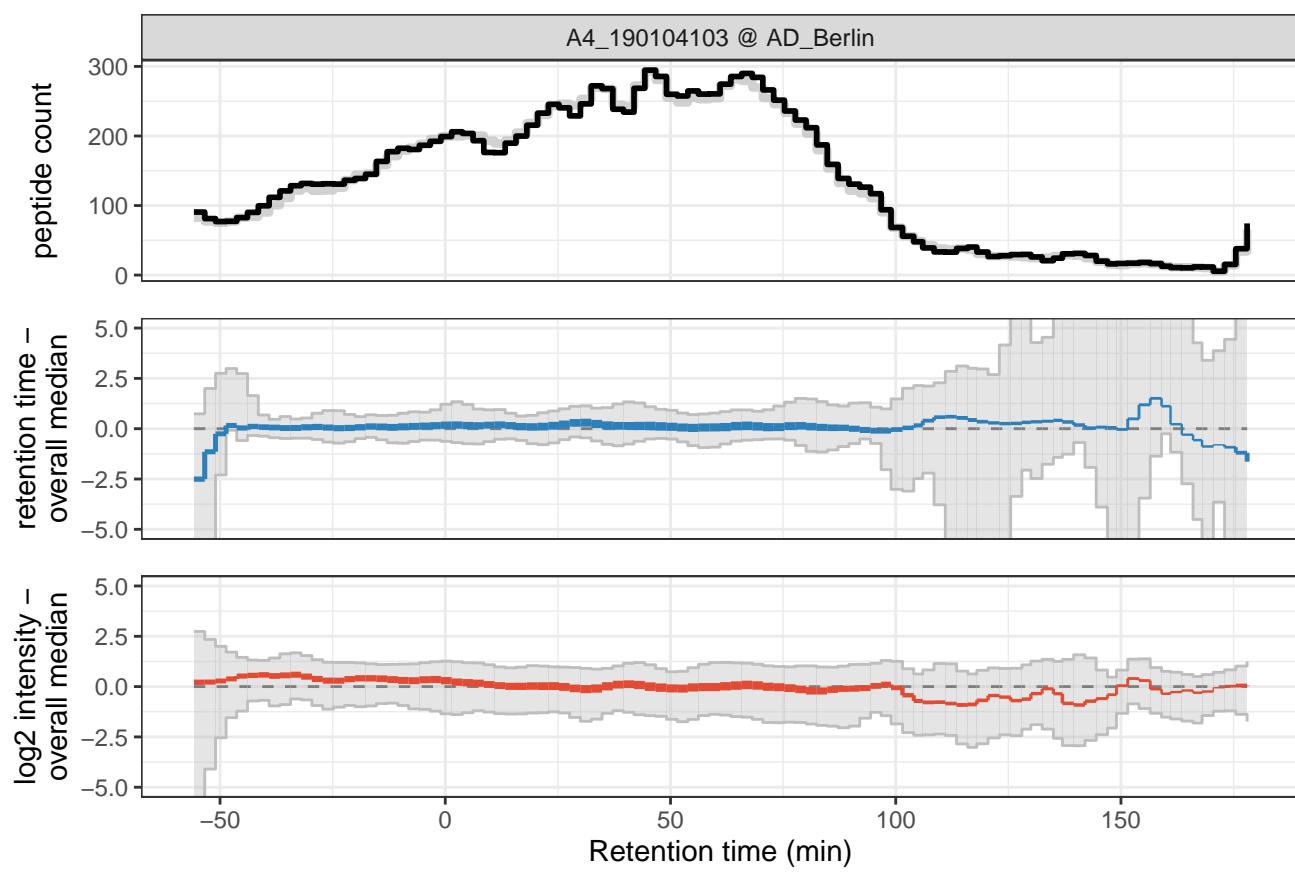
### 1.5.2 retention time local effects

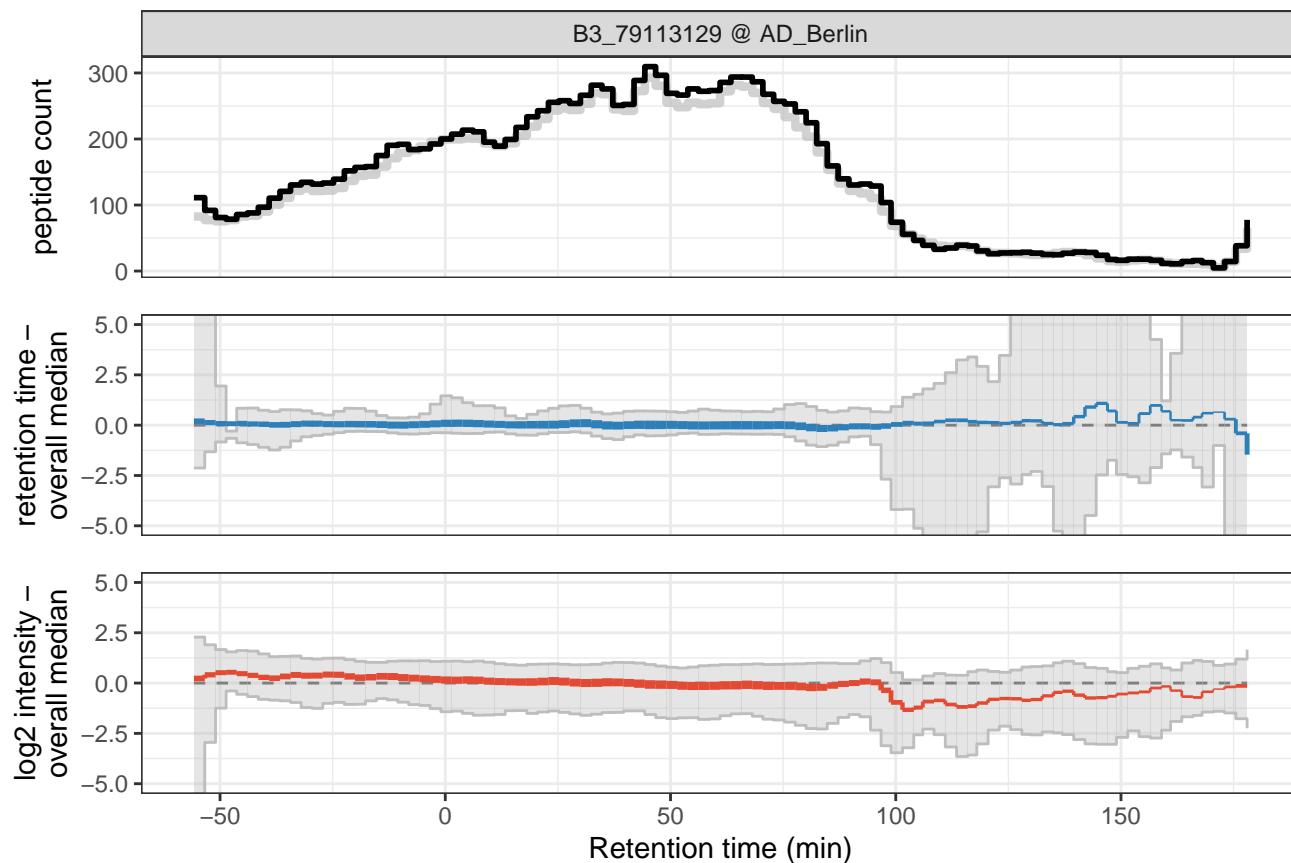
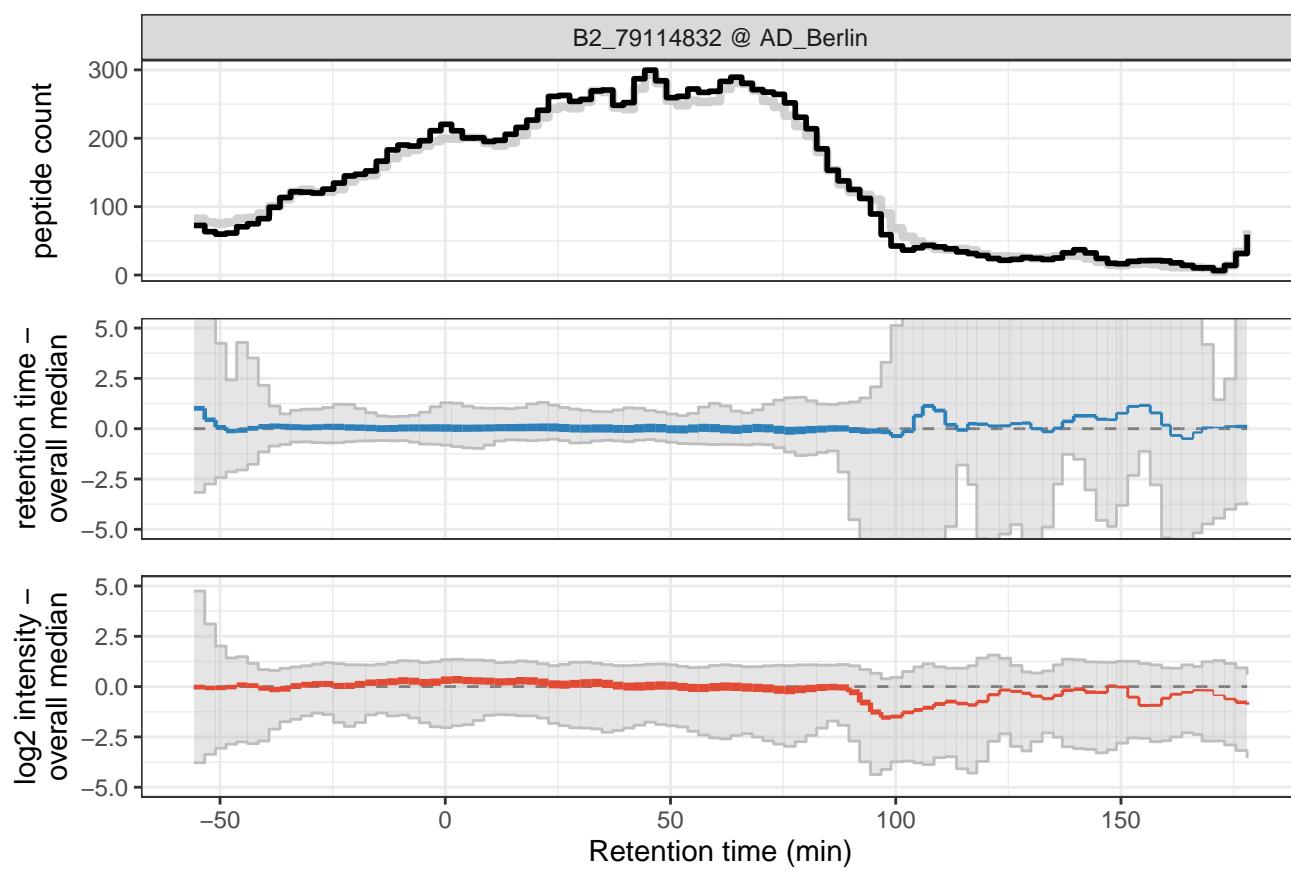
To investigate how each measurement differs from others, we visualize each sample as a 3 panel figure. First, the data is binned across the retention time dimension (x-axis). If a samples was marked as ‘exclude’ in the provided sample metadata, this is indicated in the plot title.

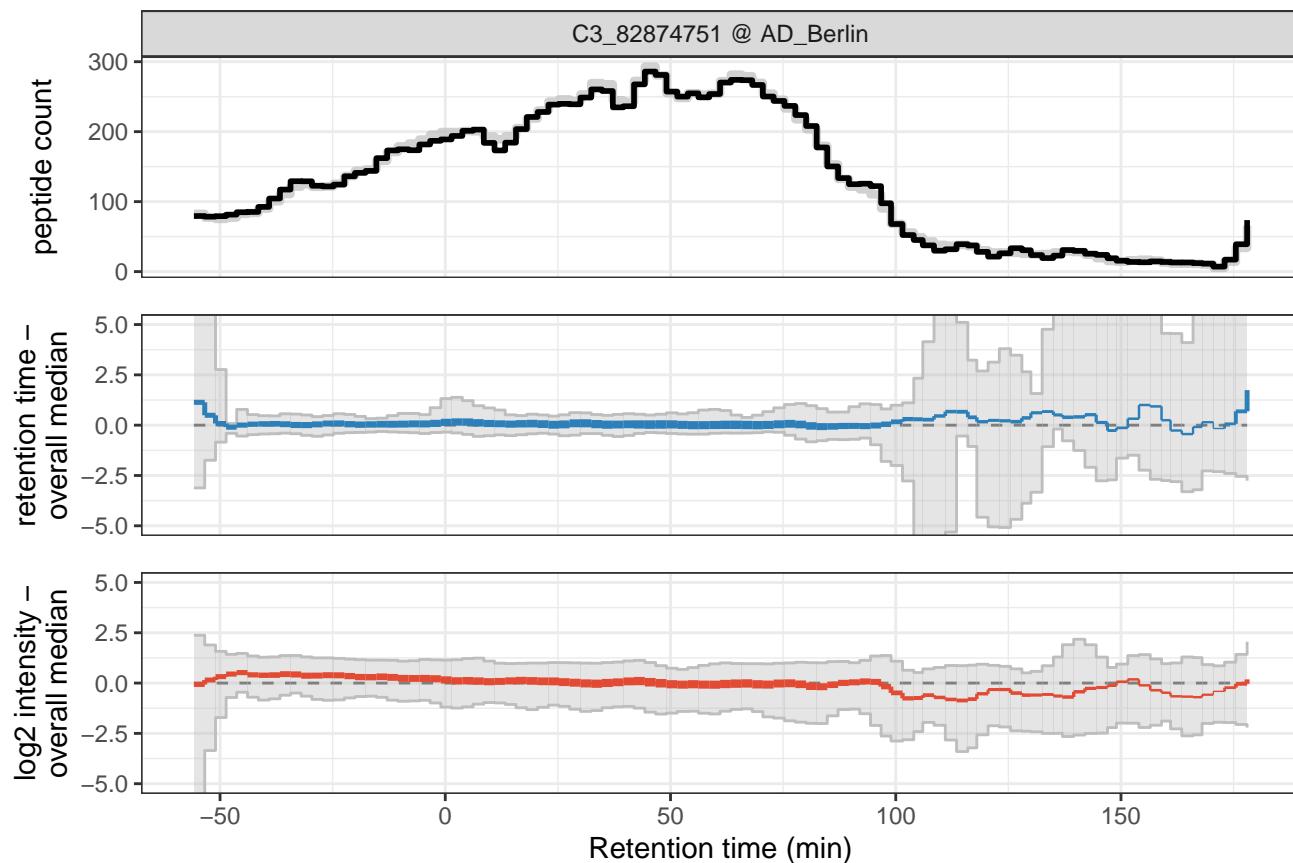
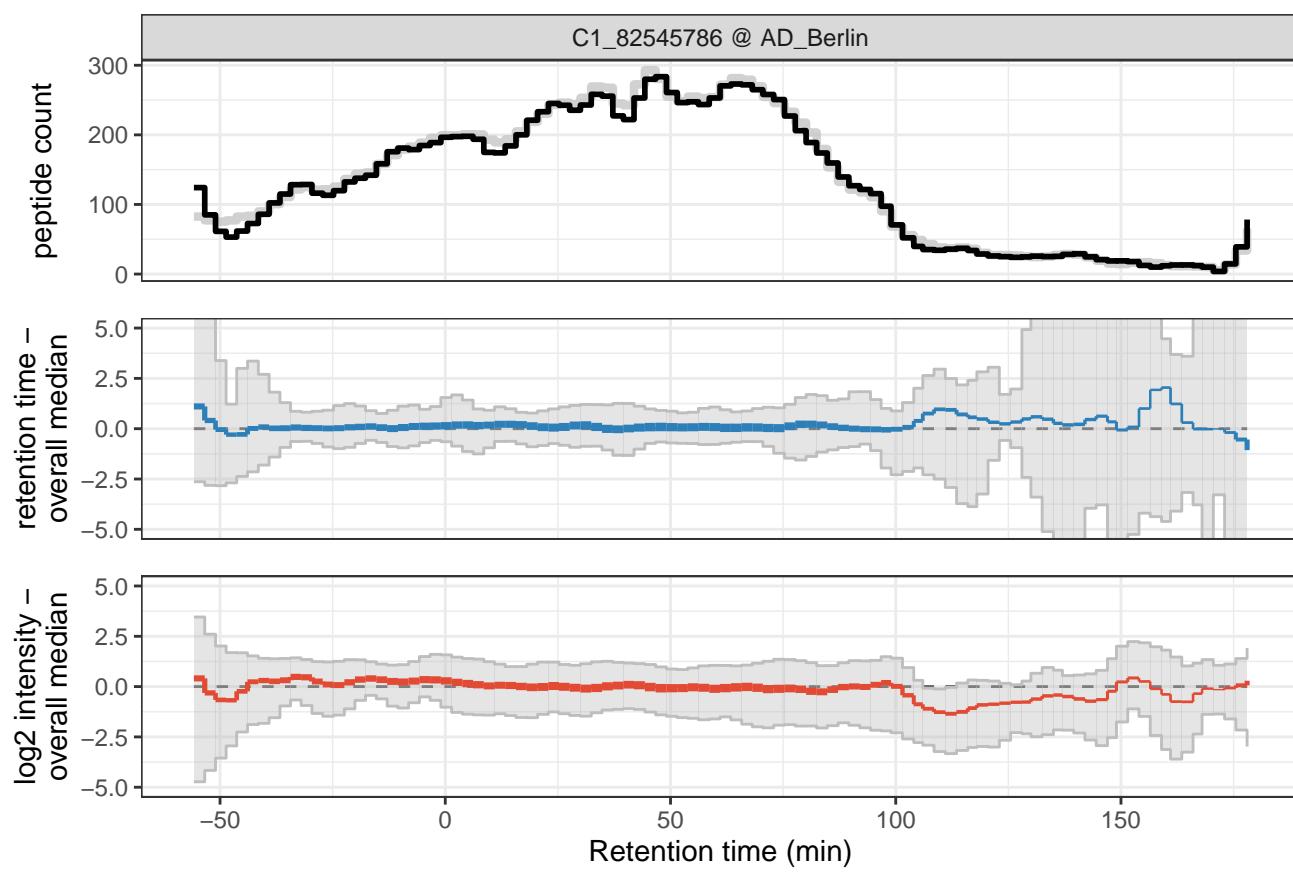
The top panel shows the number of peptides in the input data, e.g. as recognized by the software that generated input for this pipeline, over time (black line). For reference, the grey line shows the median amount over all samples (note; if this is the exact same in all samples, the grey line may not be visible as it falls behind the black line).

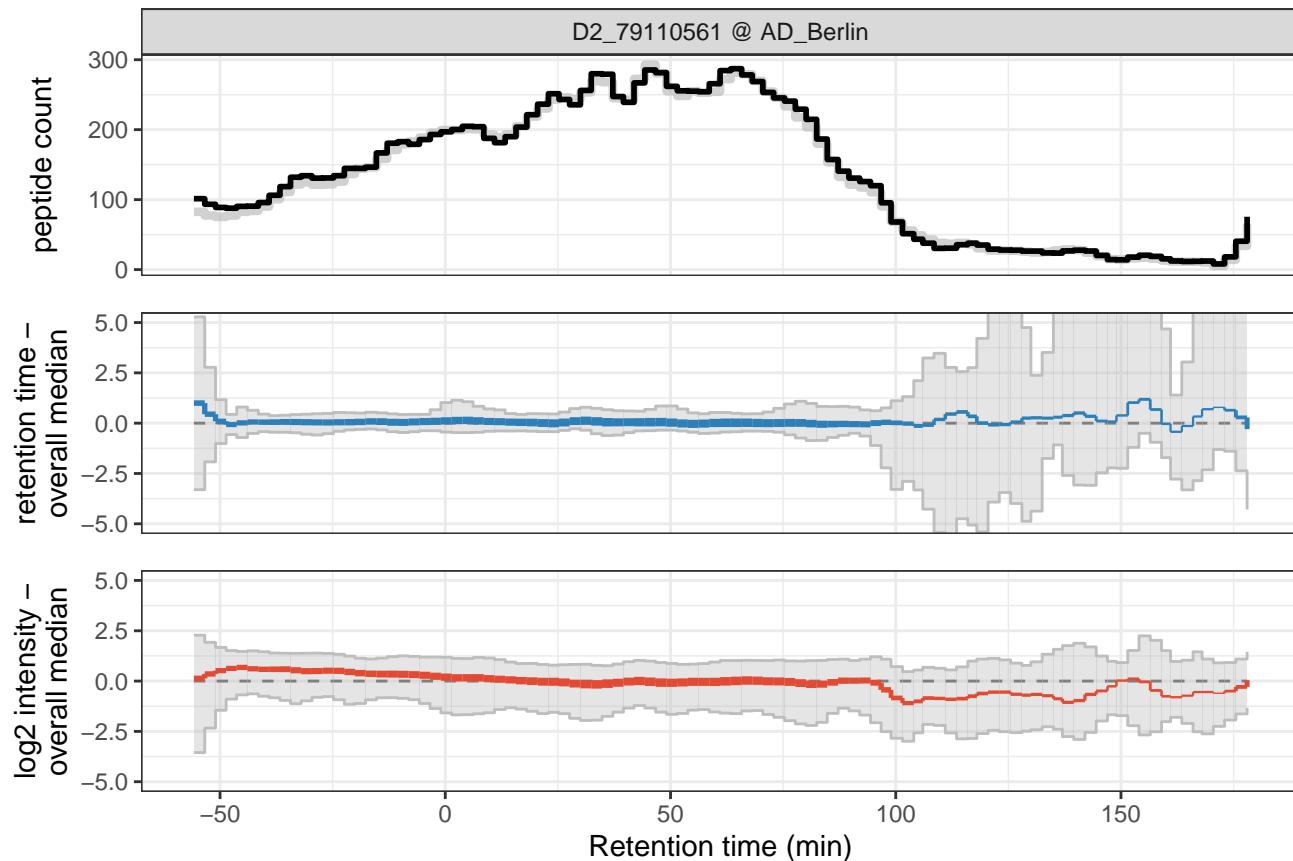
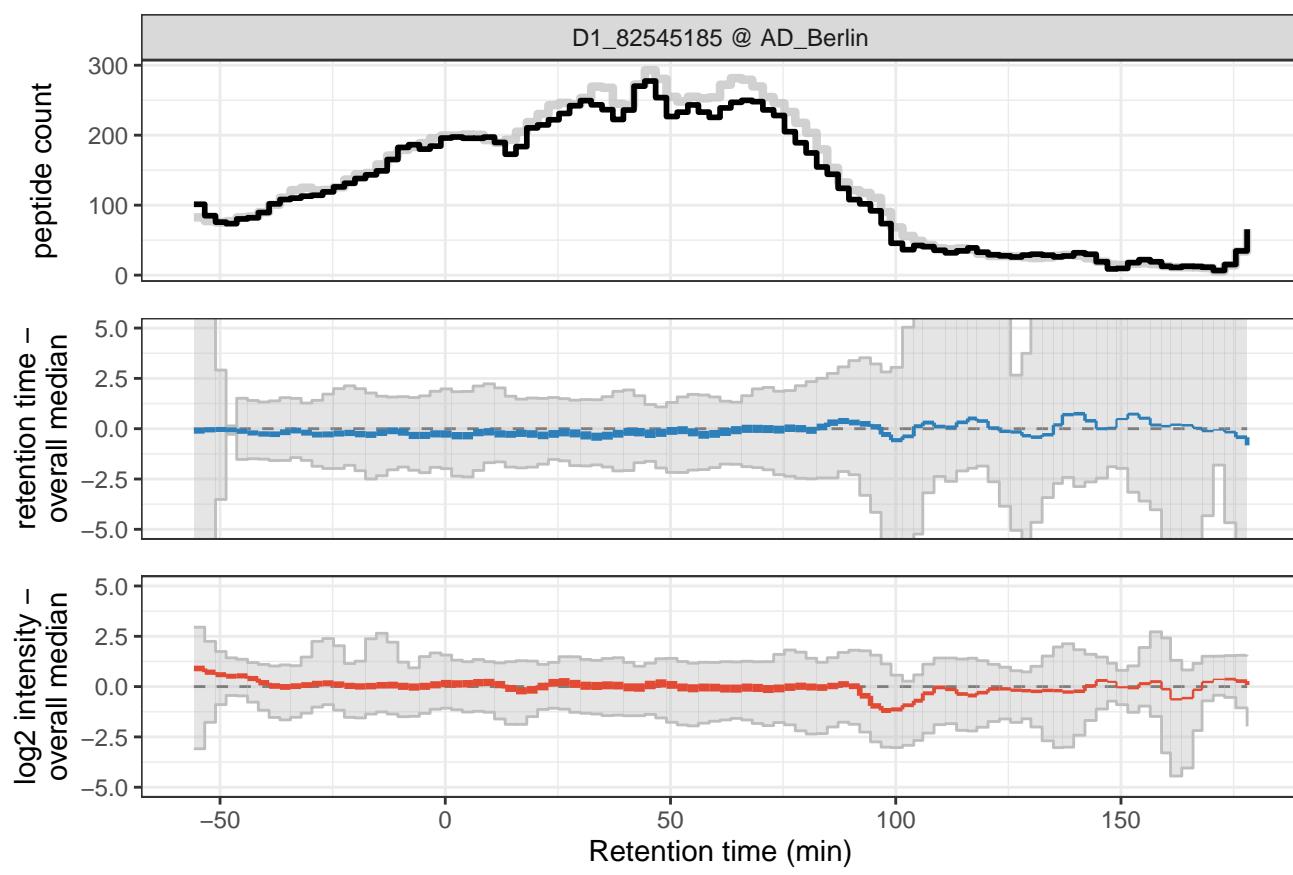
The middle panel indicates whether peptide retention times deviate from their median over all samples (blue line). The grey area depicts the 5% and 95% quantiles, respectively. The line width corresponds to the number of peptides eluting at that time (data from first panel). Analogously, the bottom panel shows the deviation in peptide abundance as compared to the median over all samples (red line).

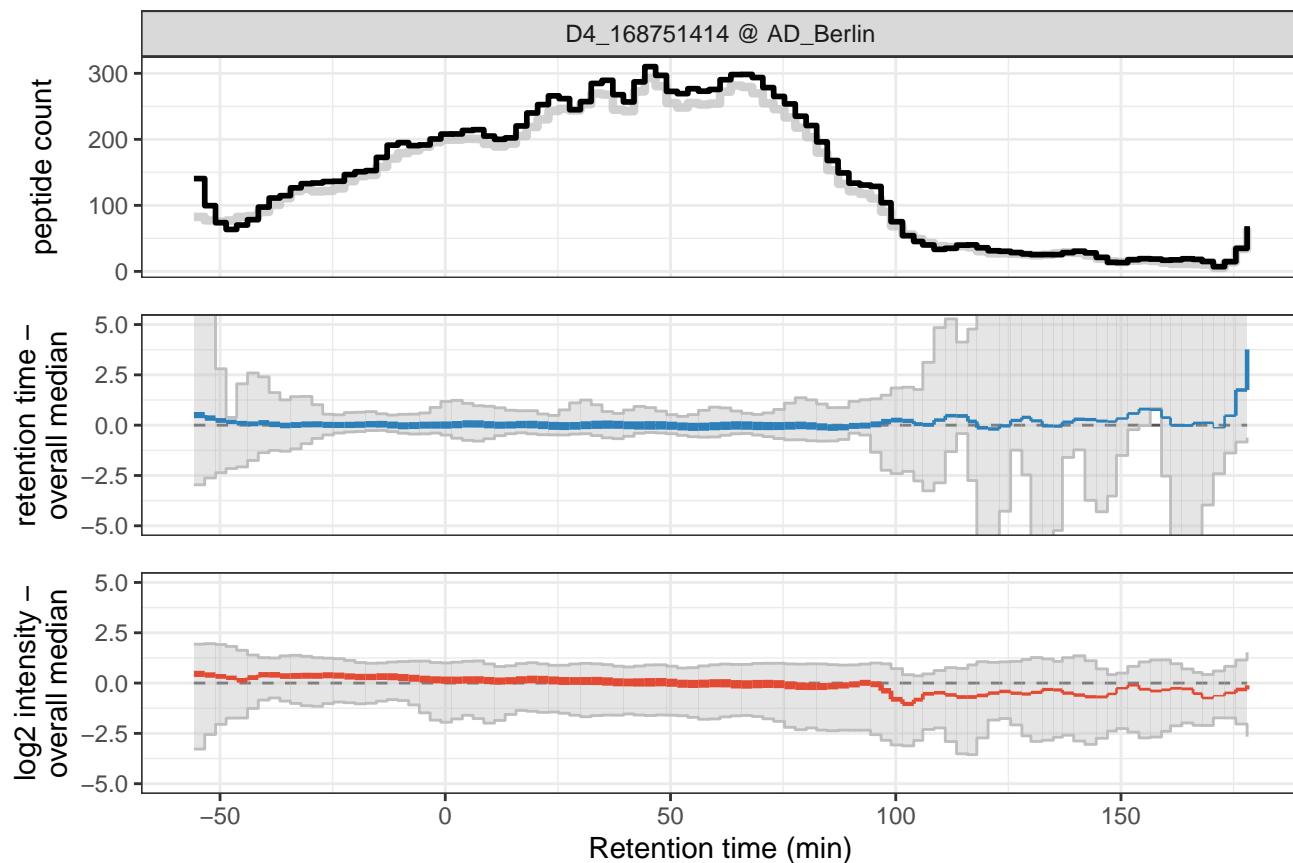
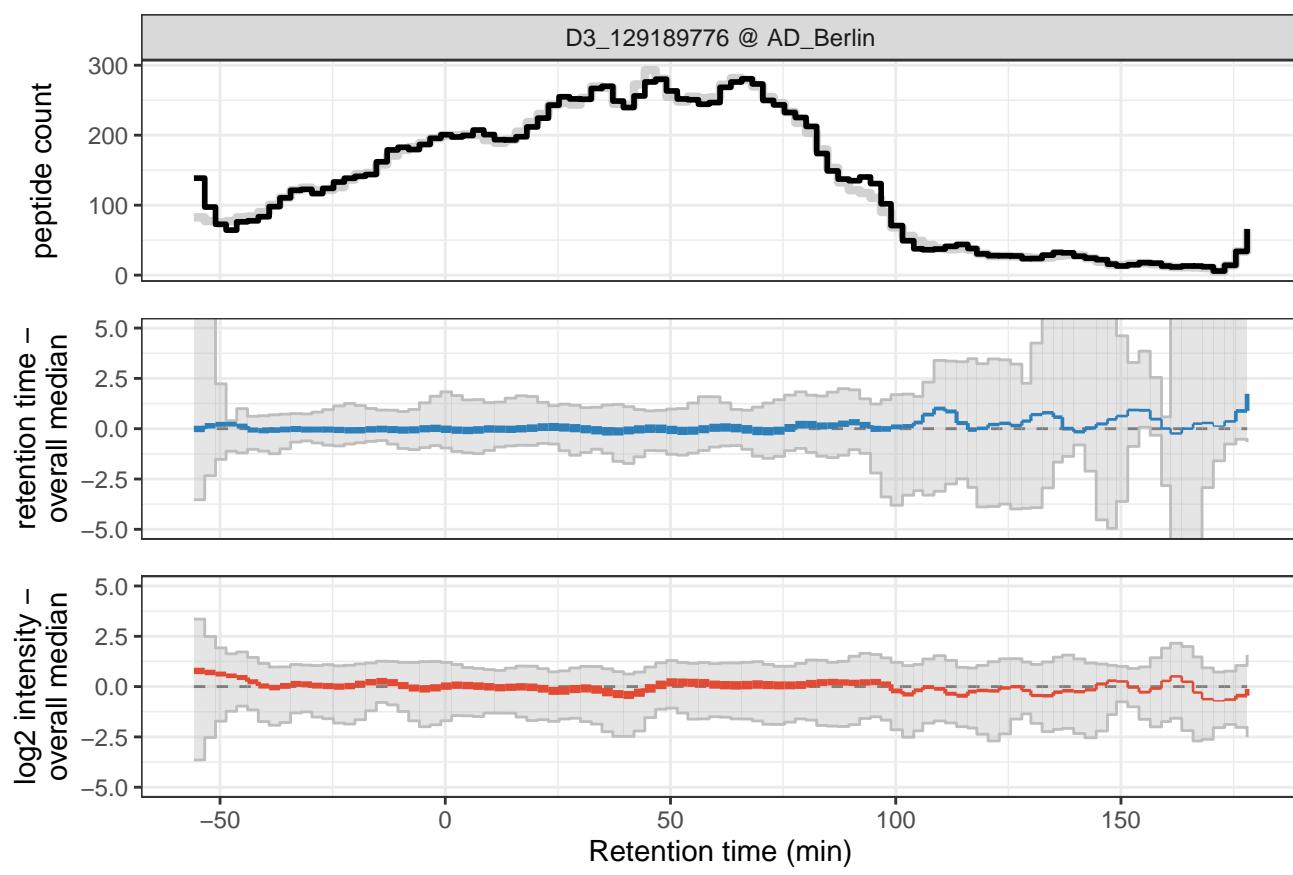


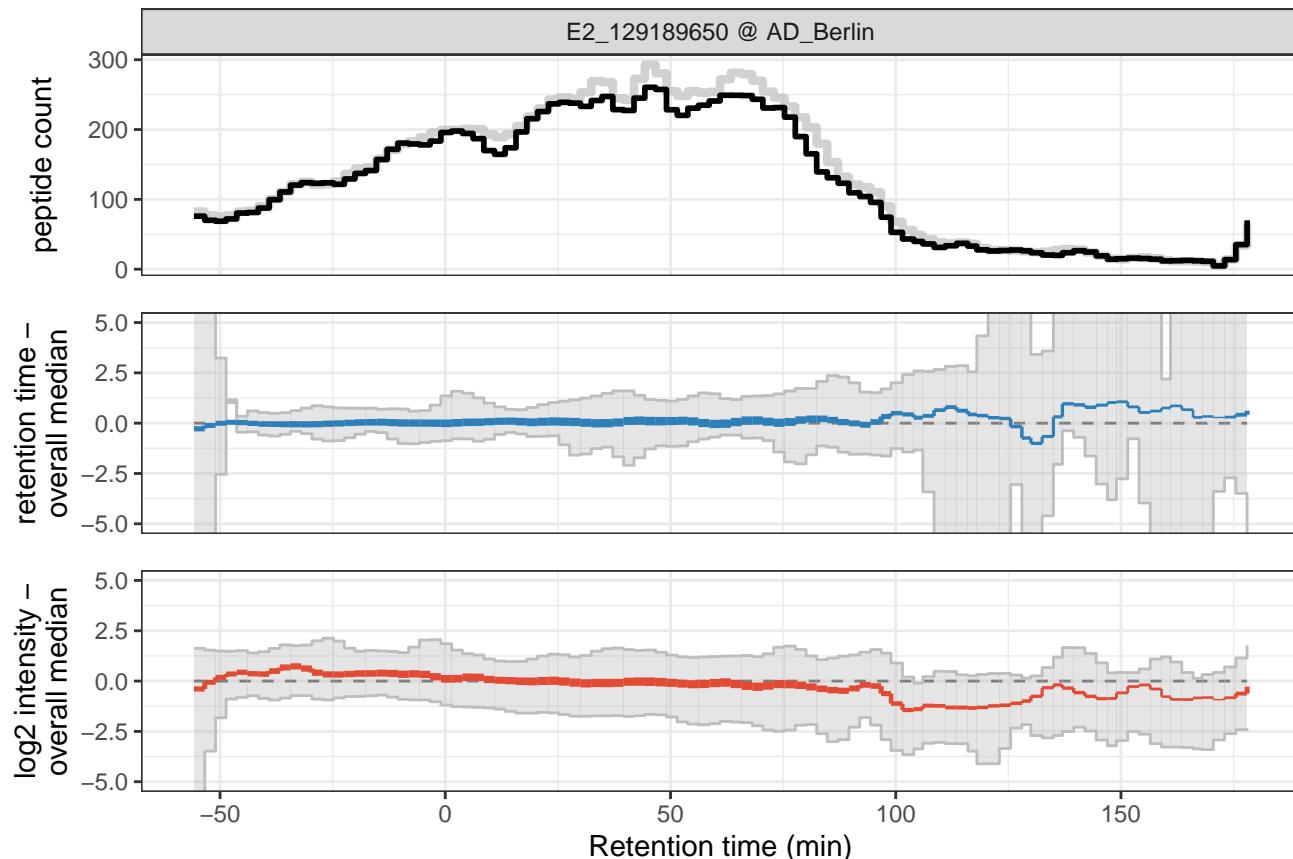
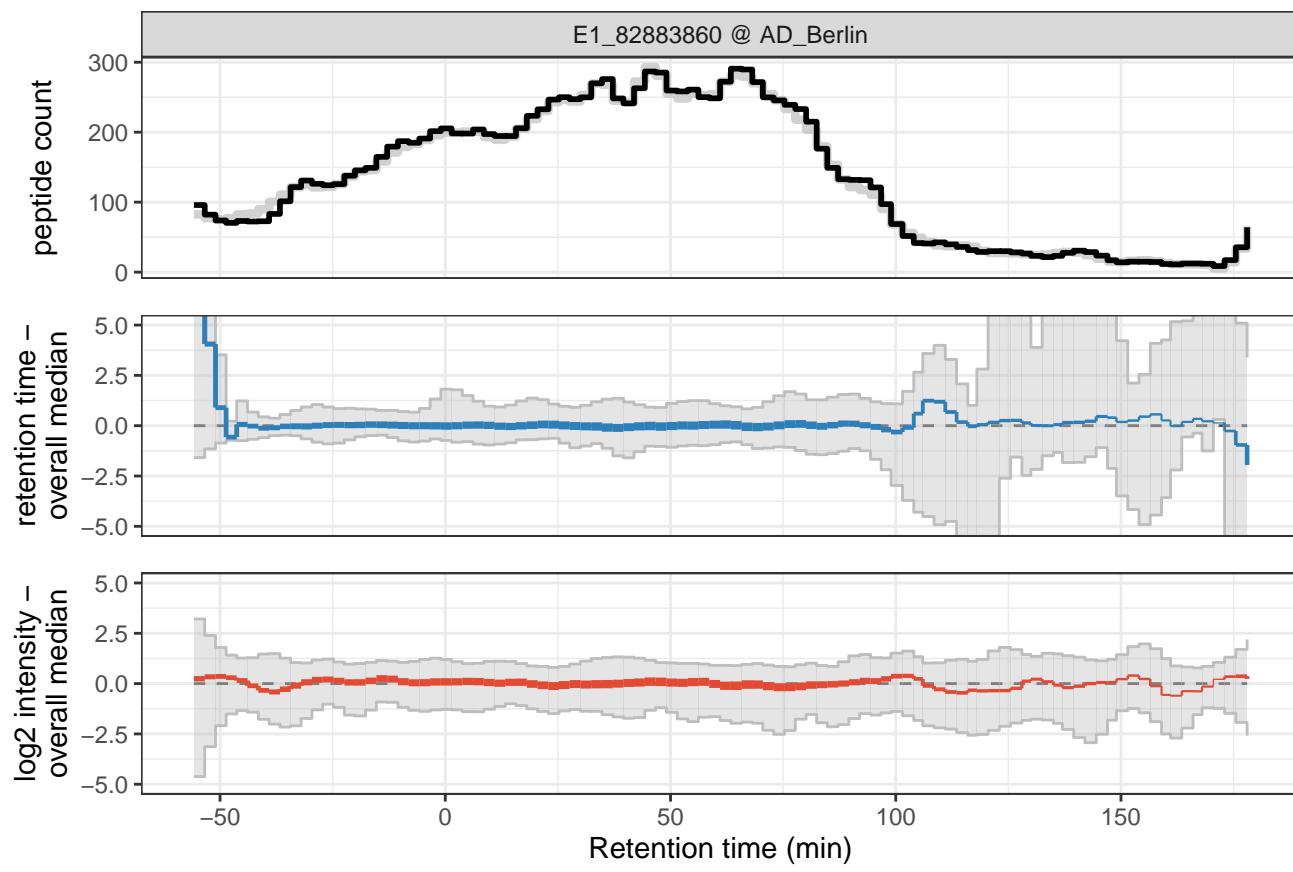


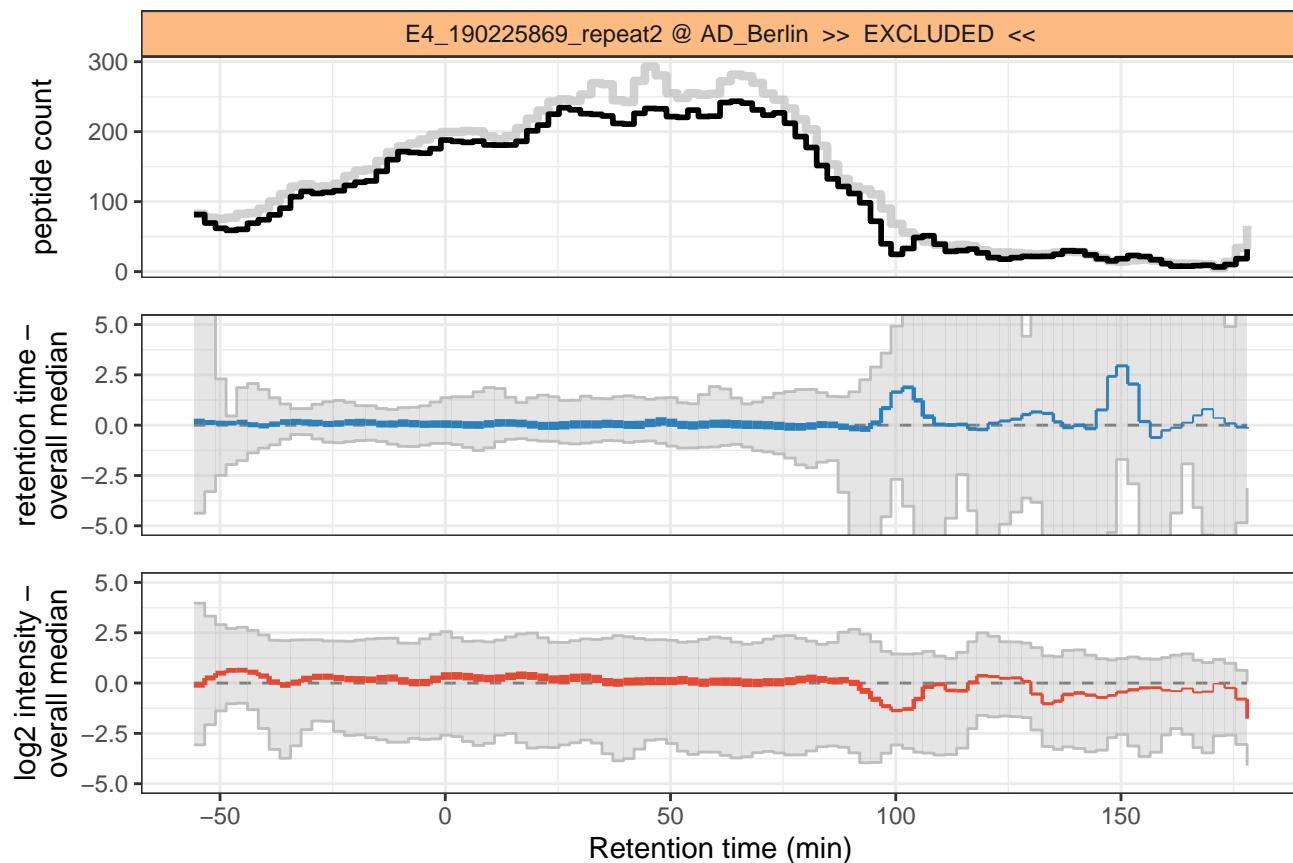
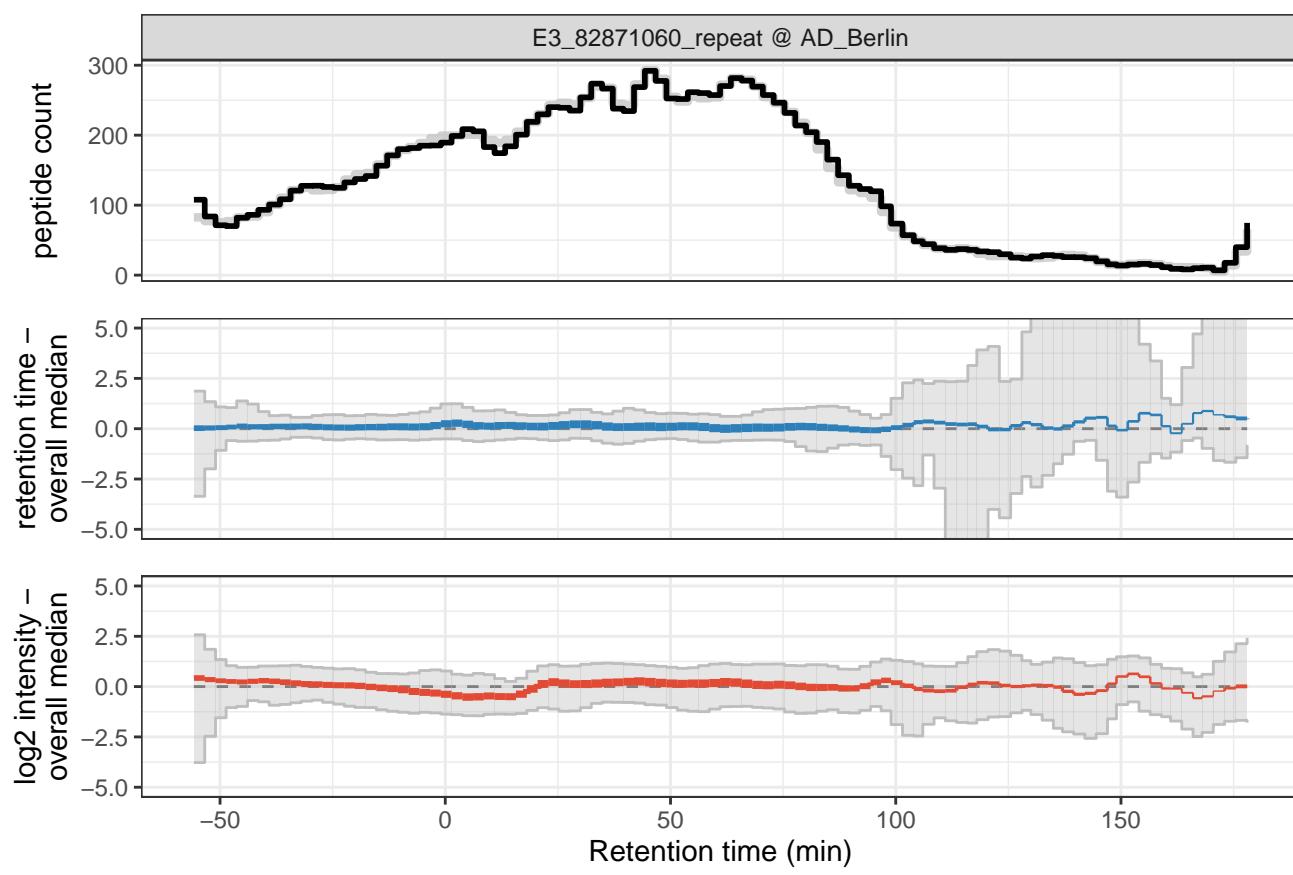


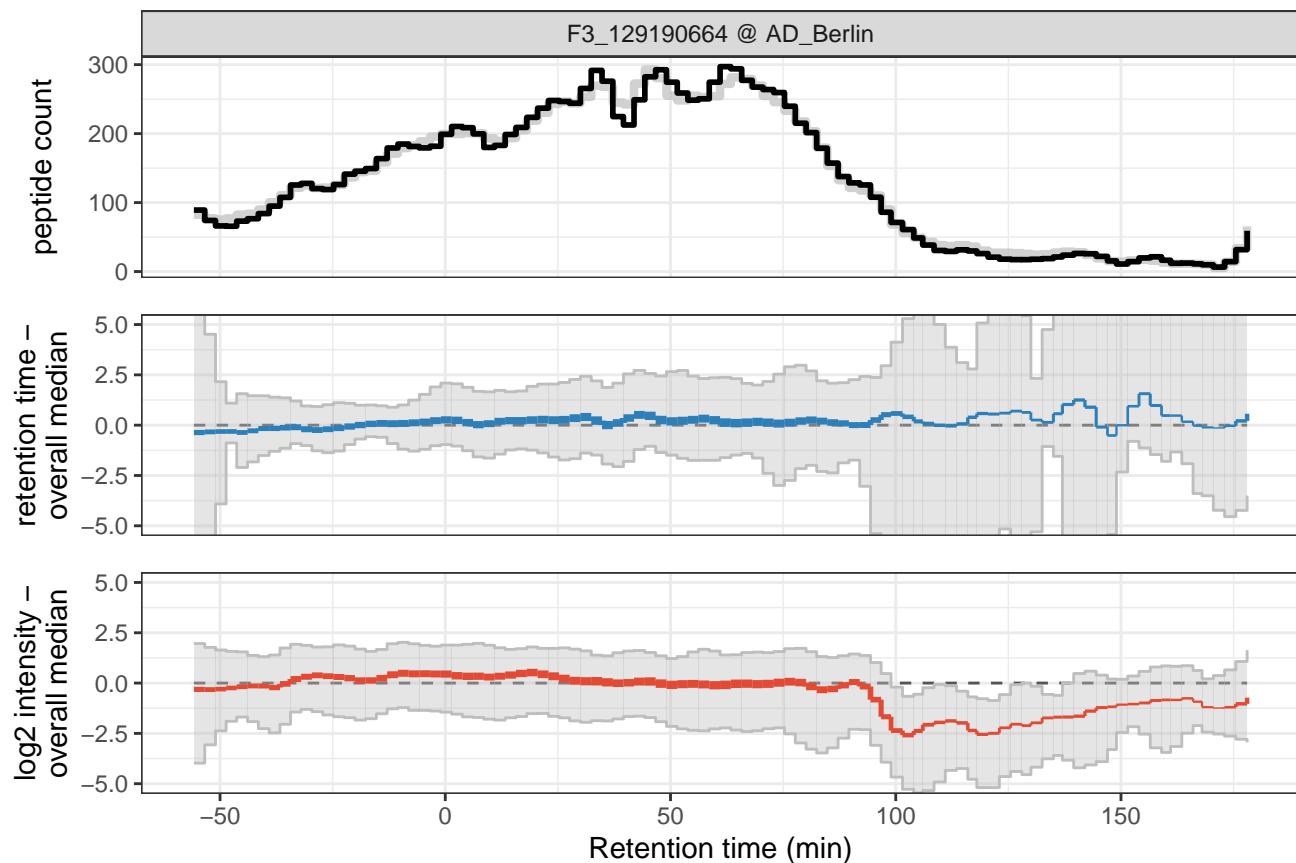
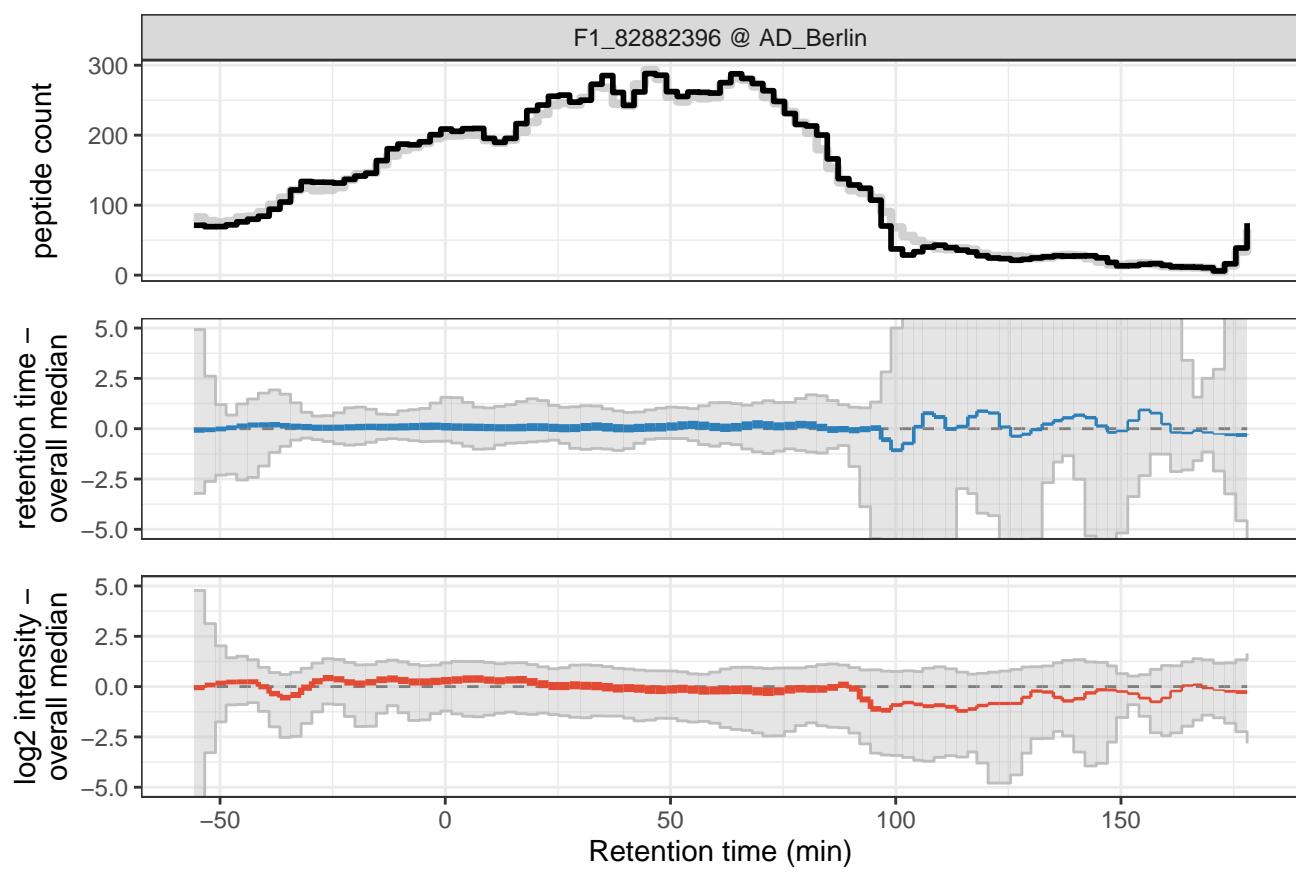


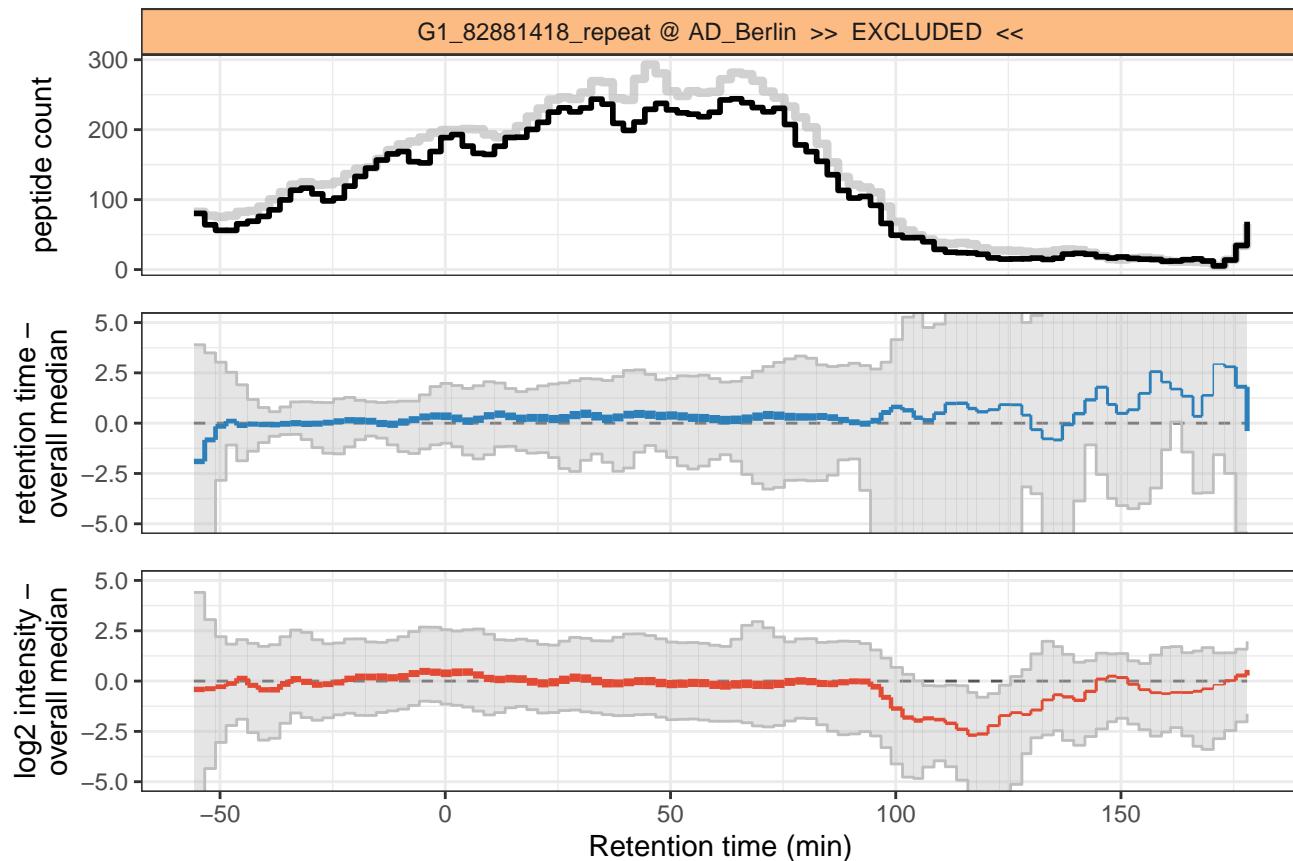
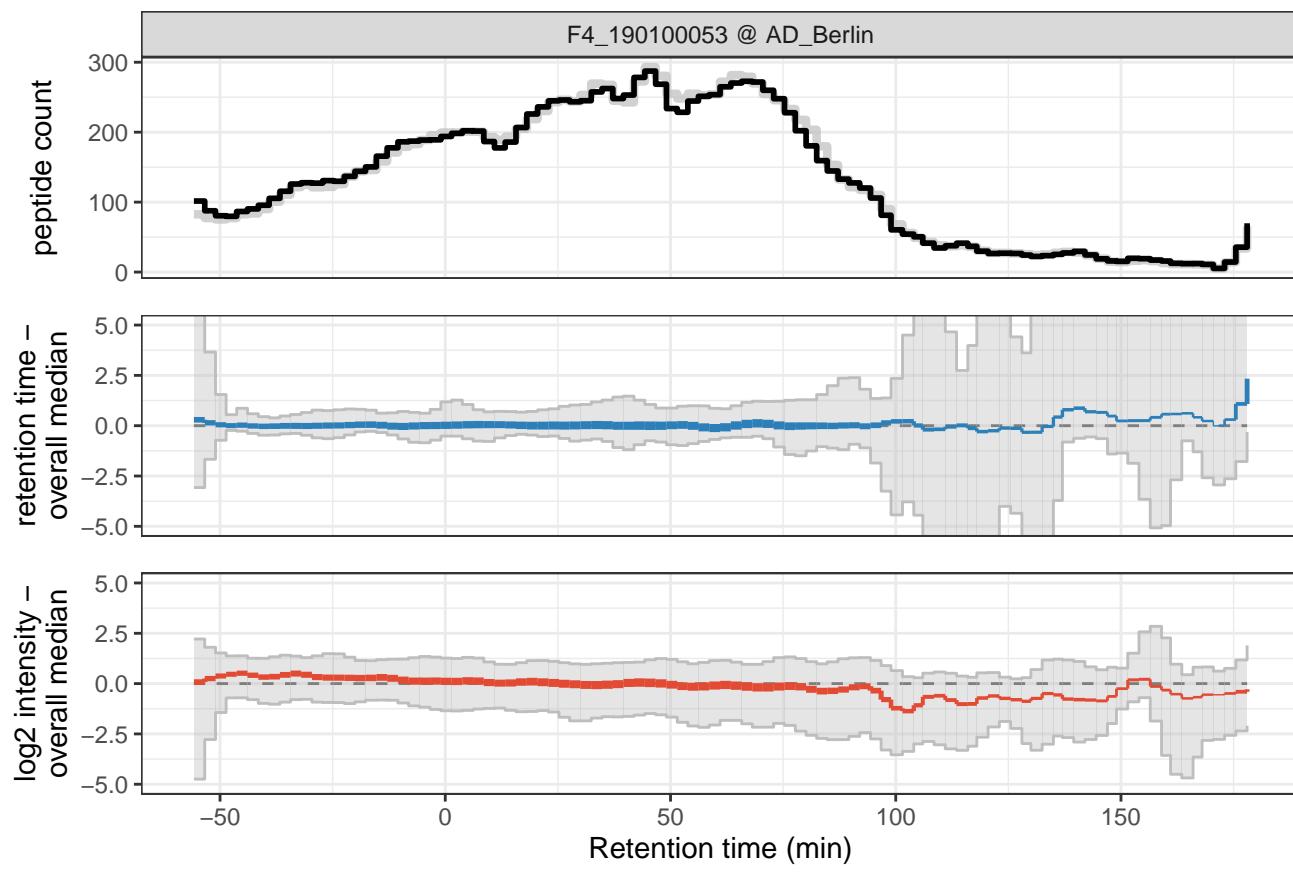


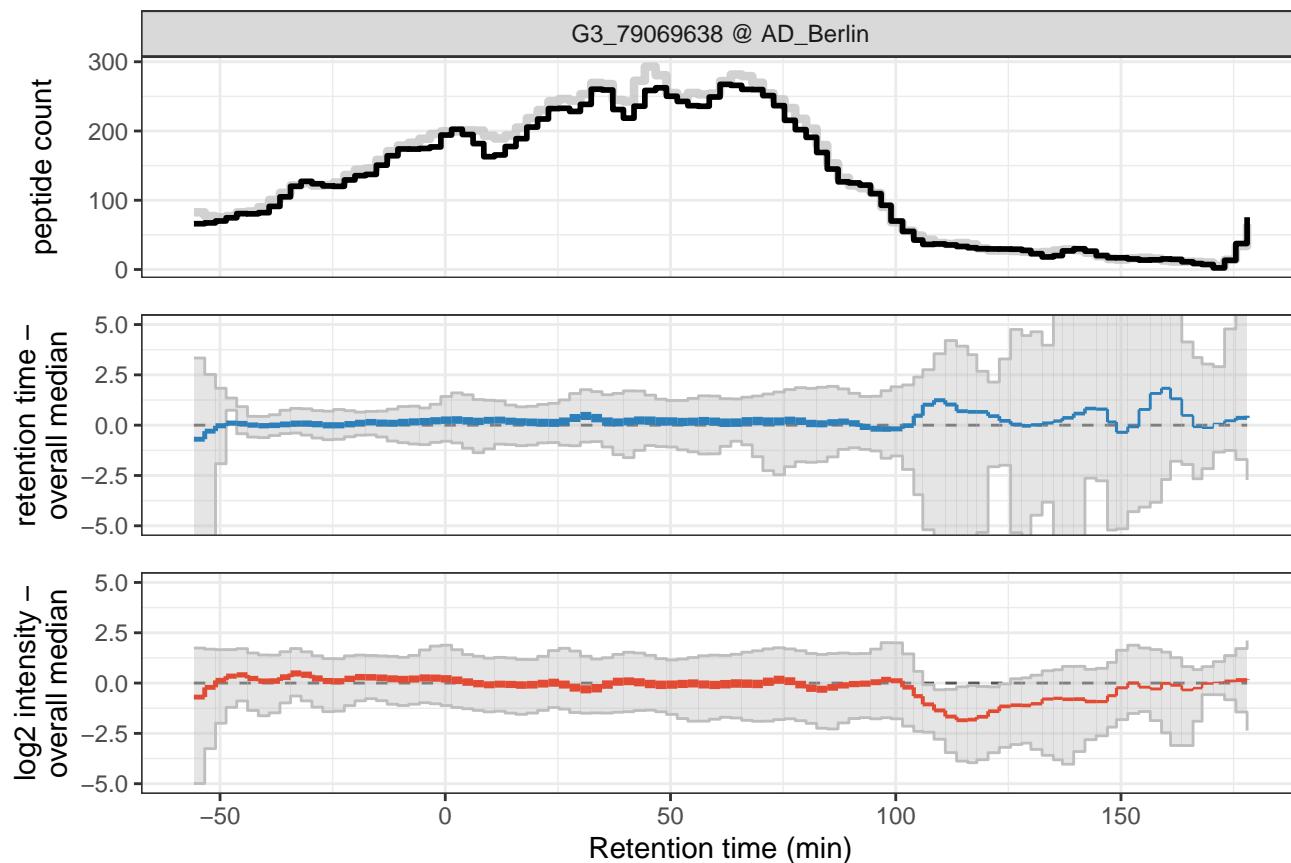
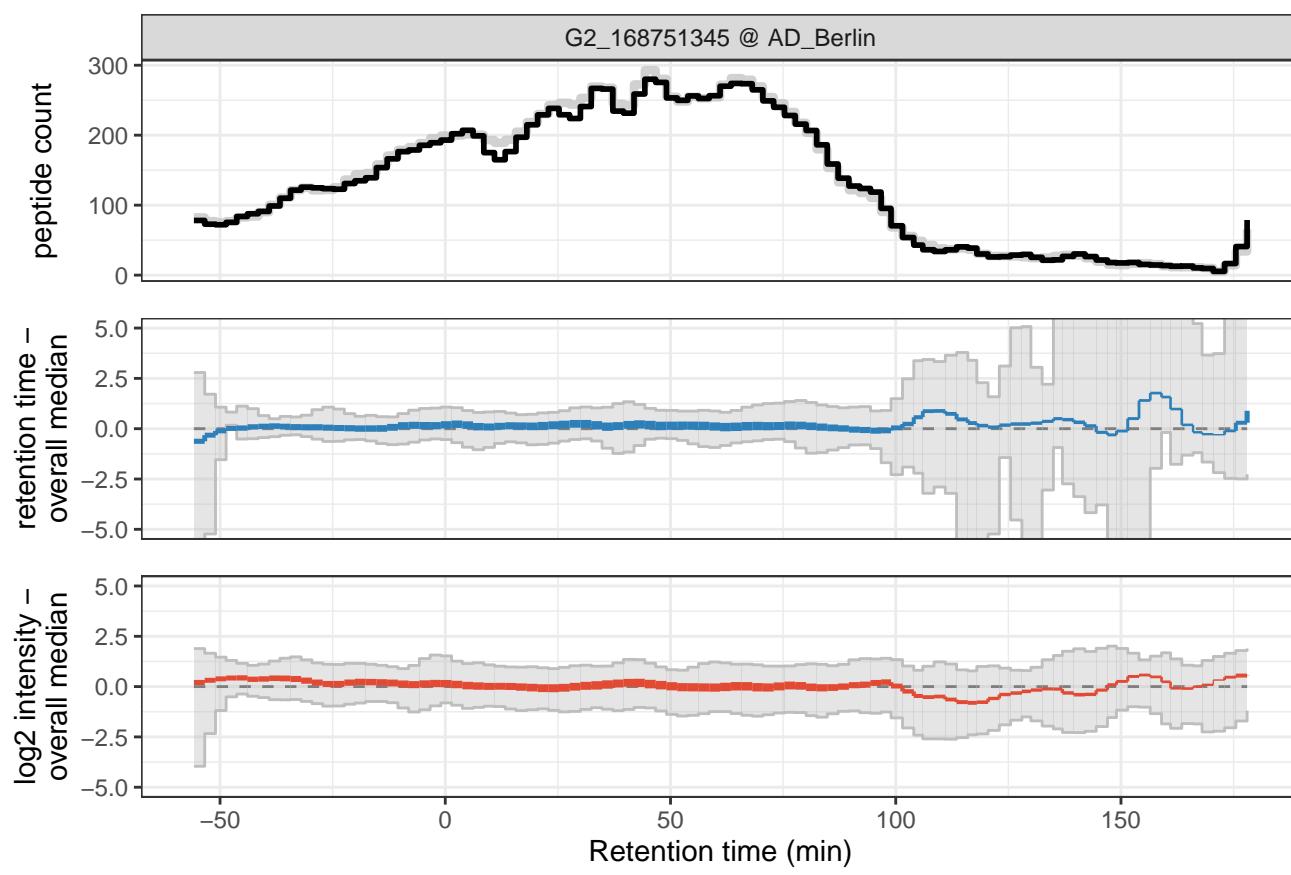


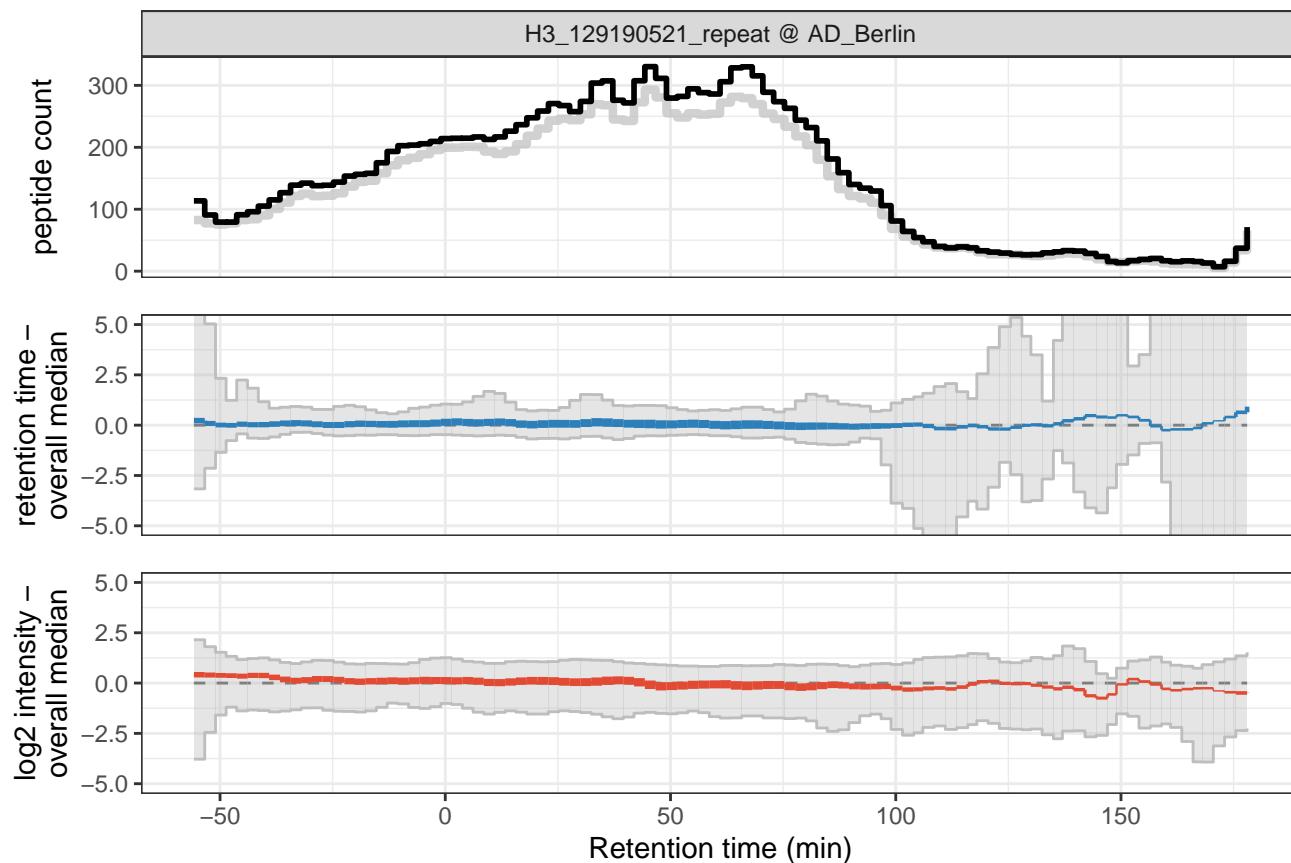
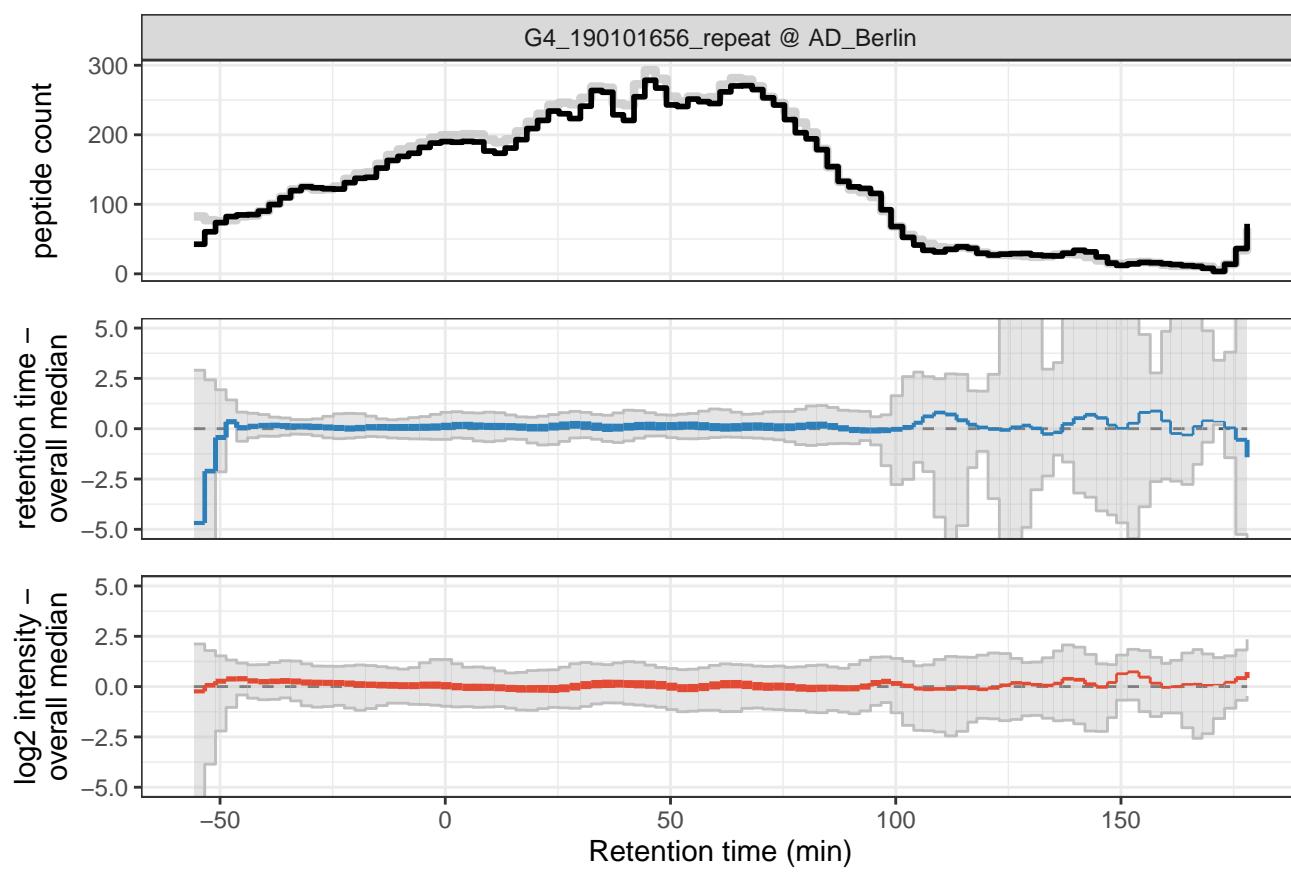


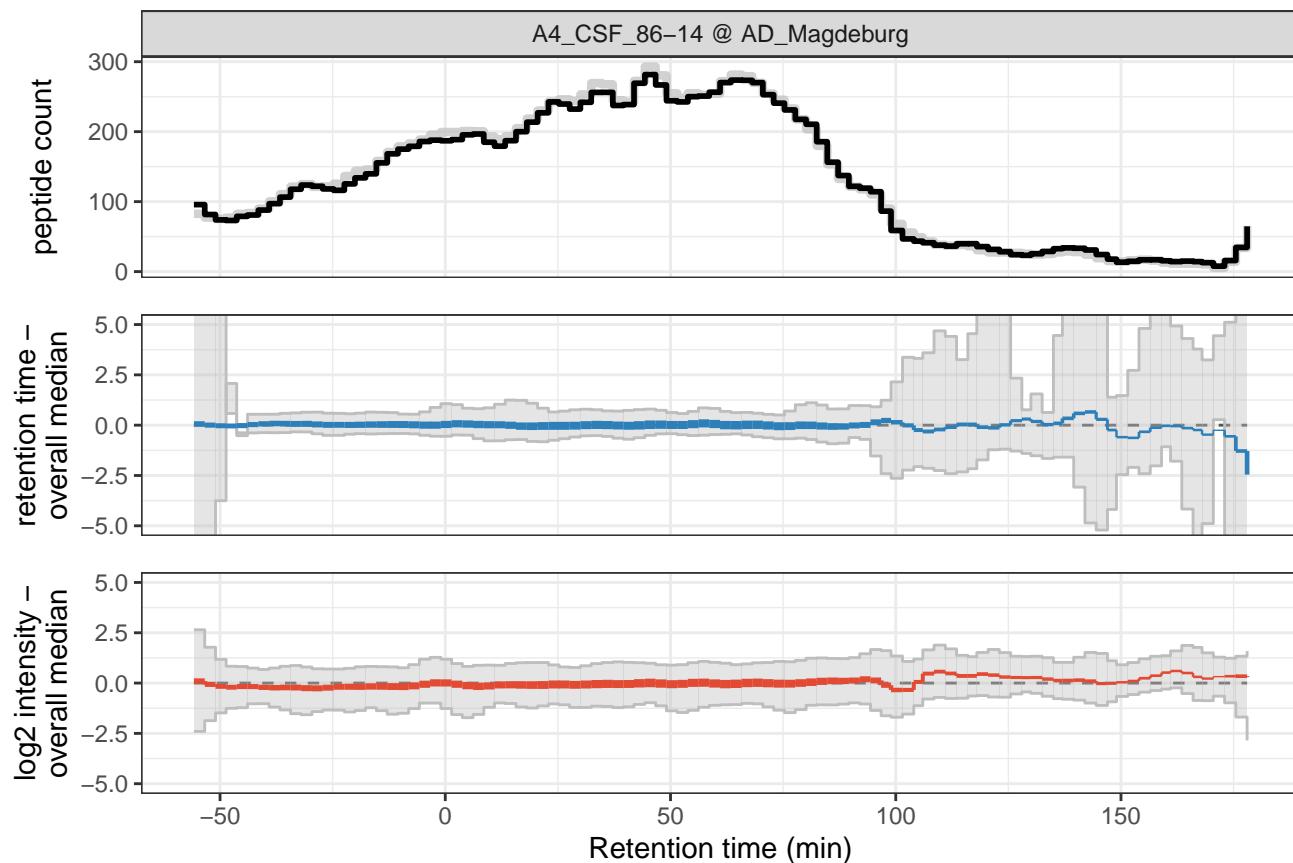
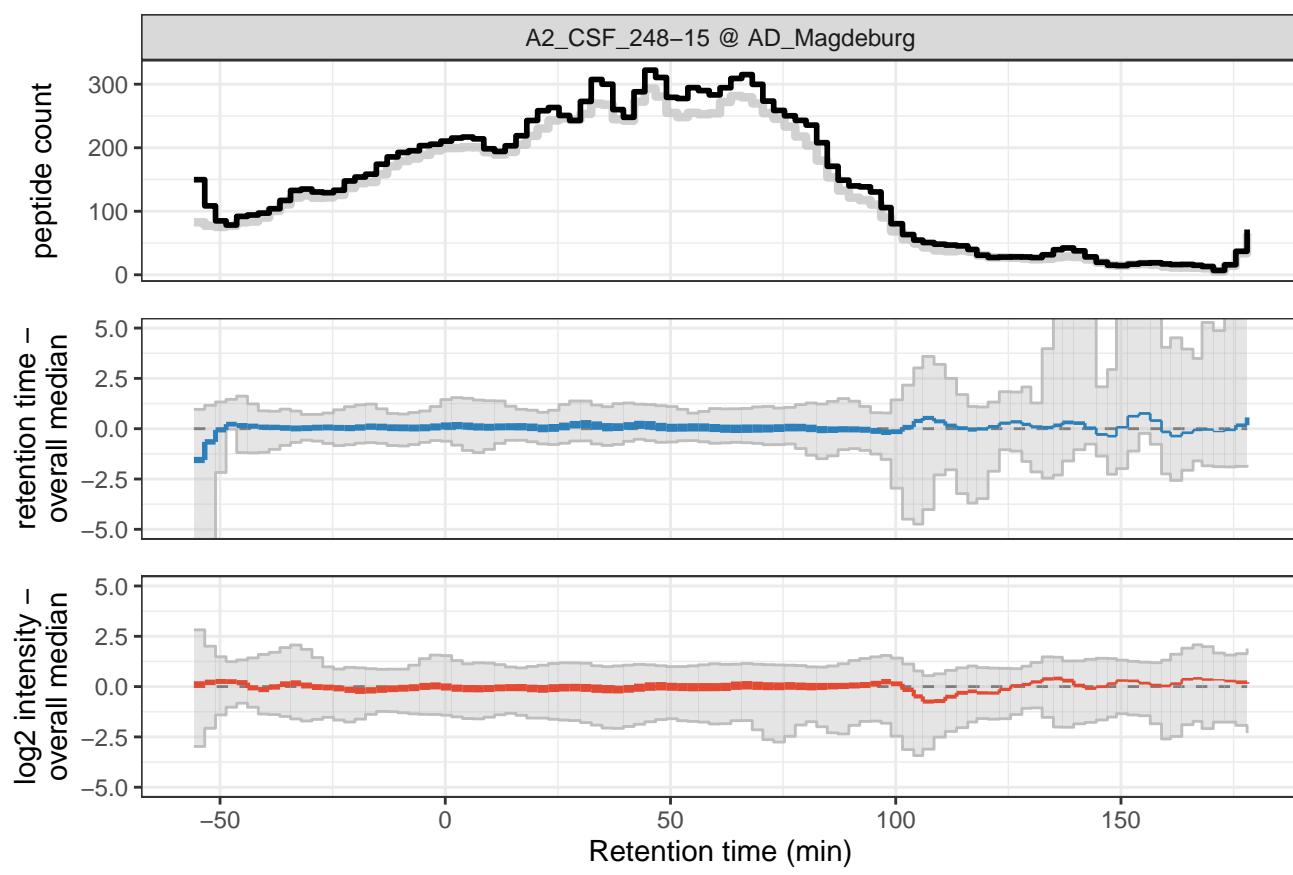


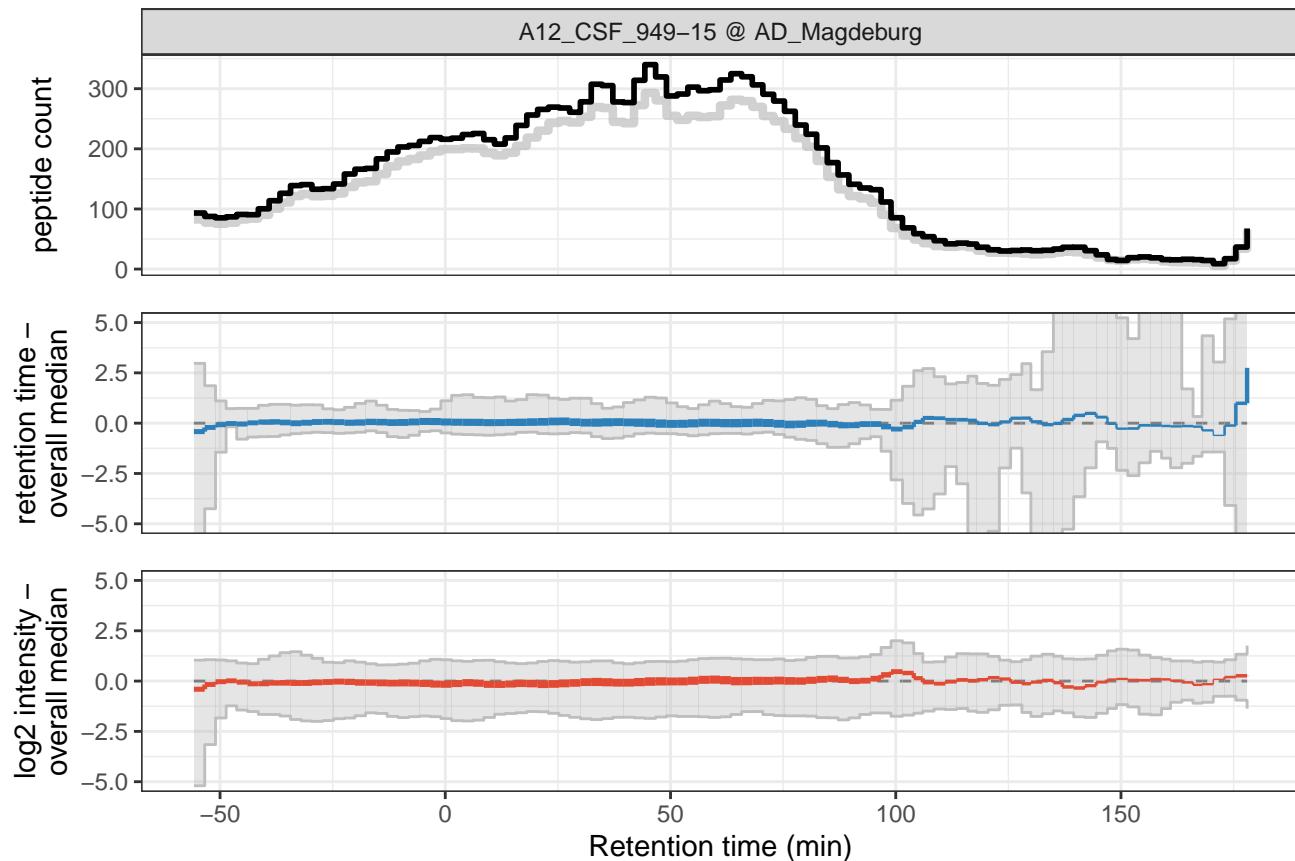
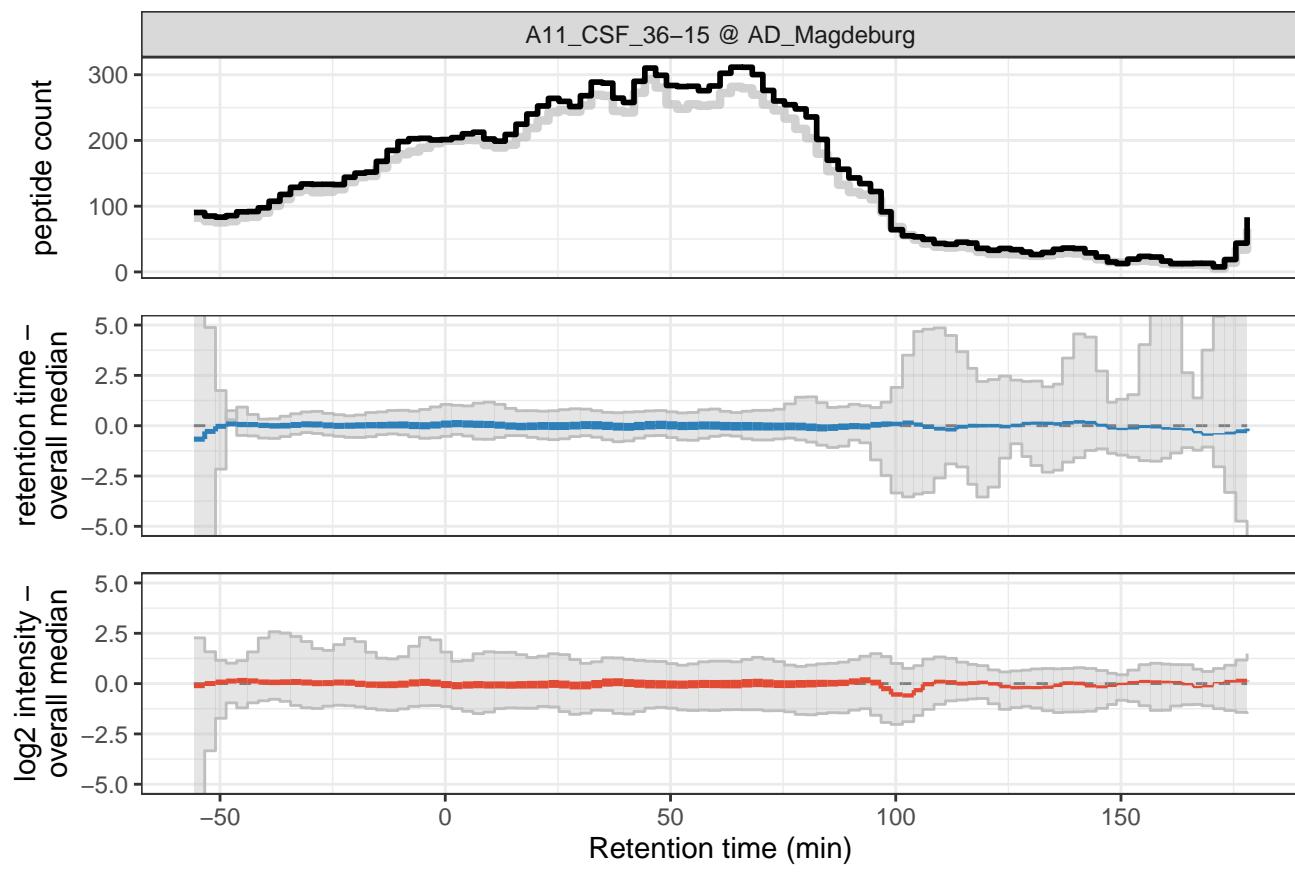


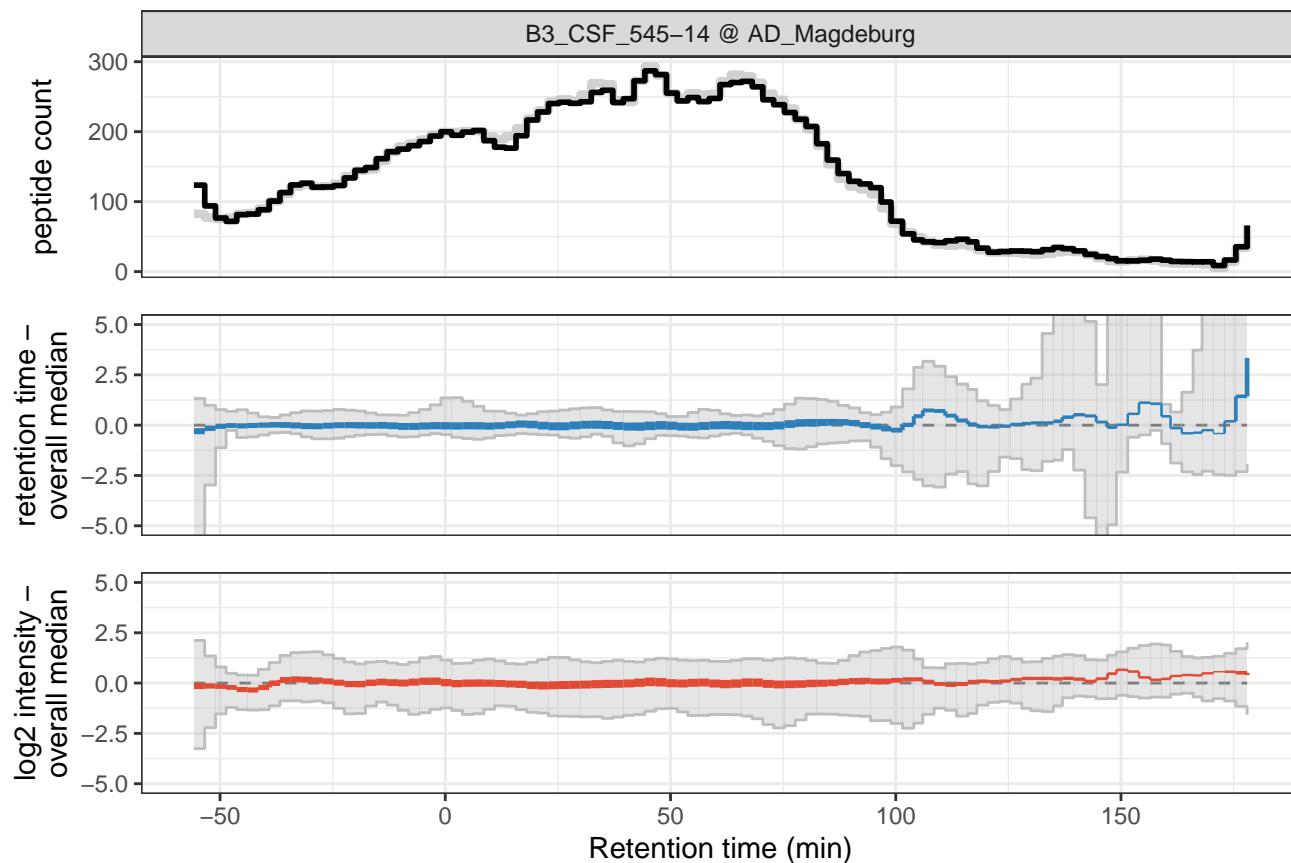
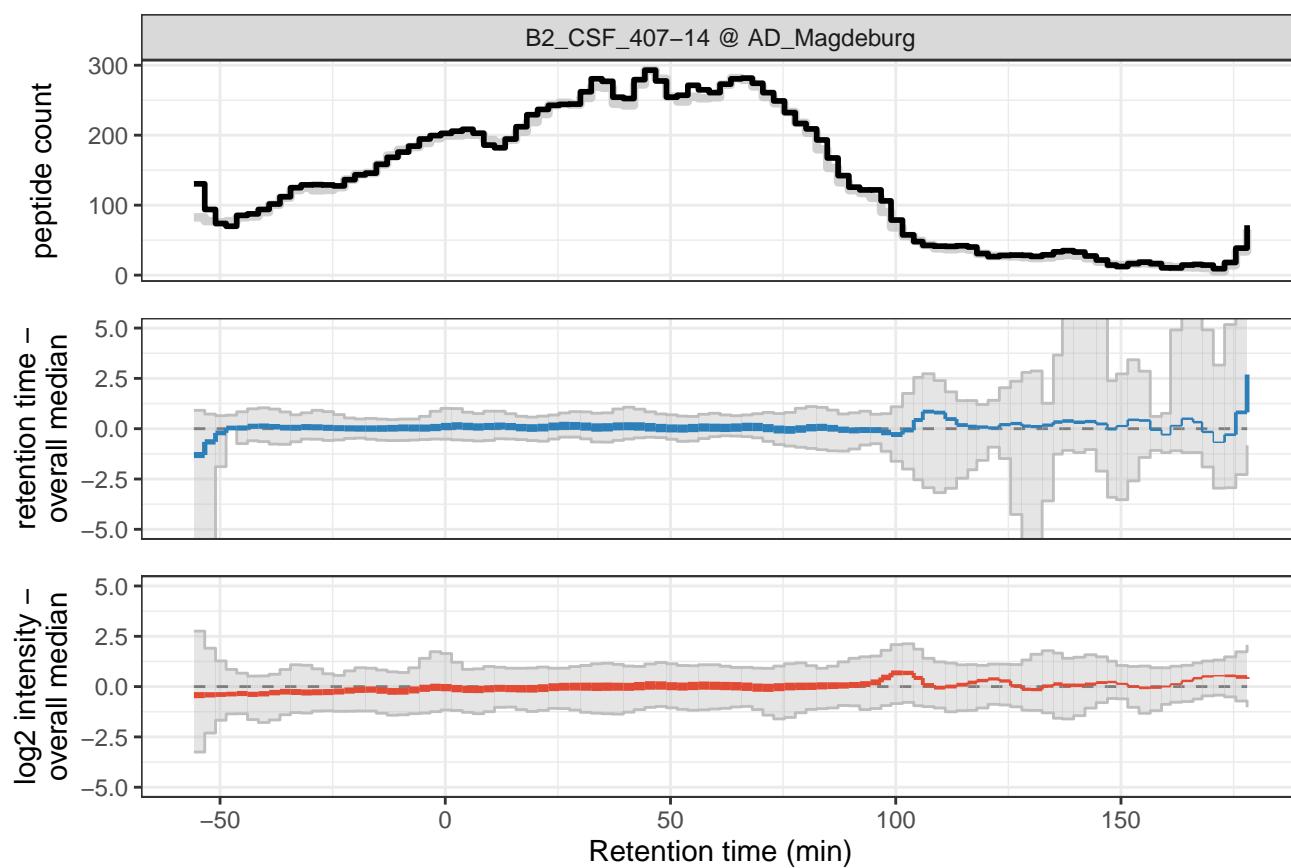


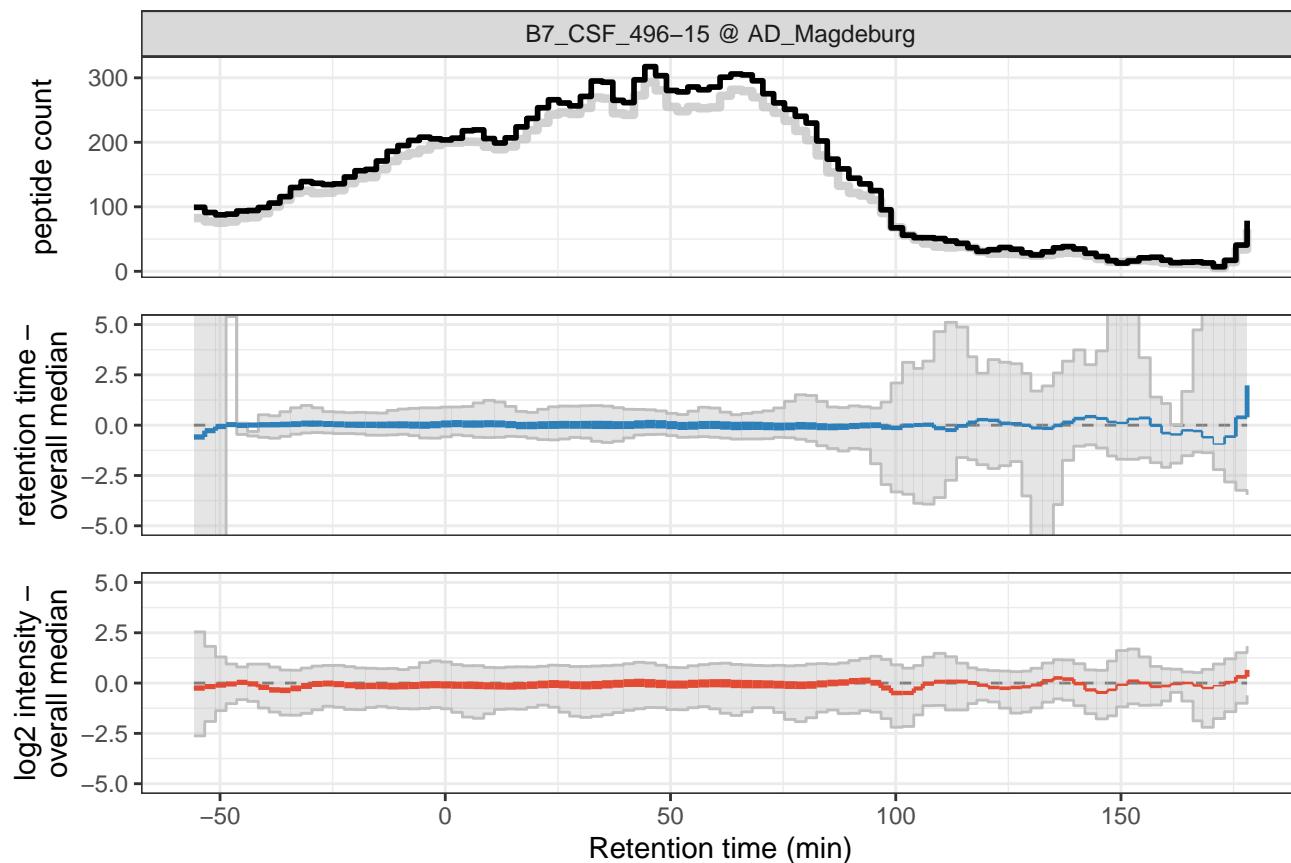
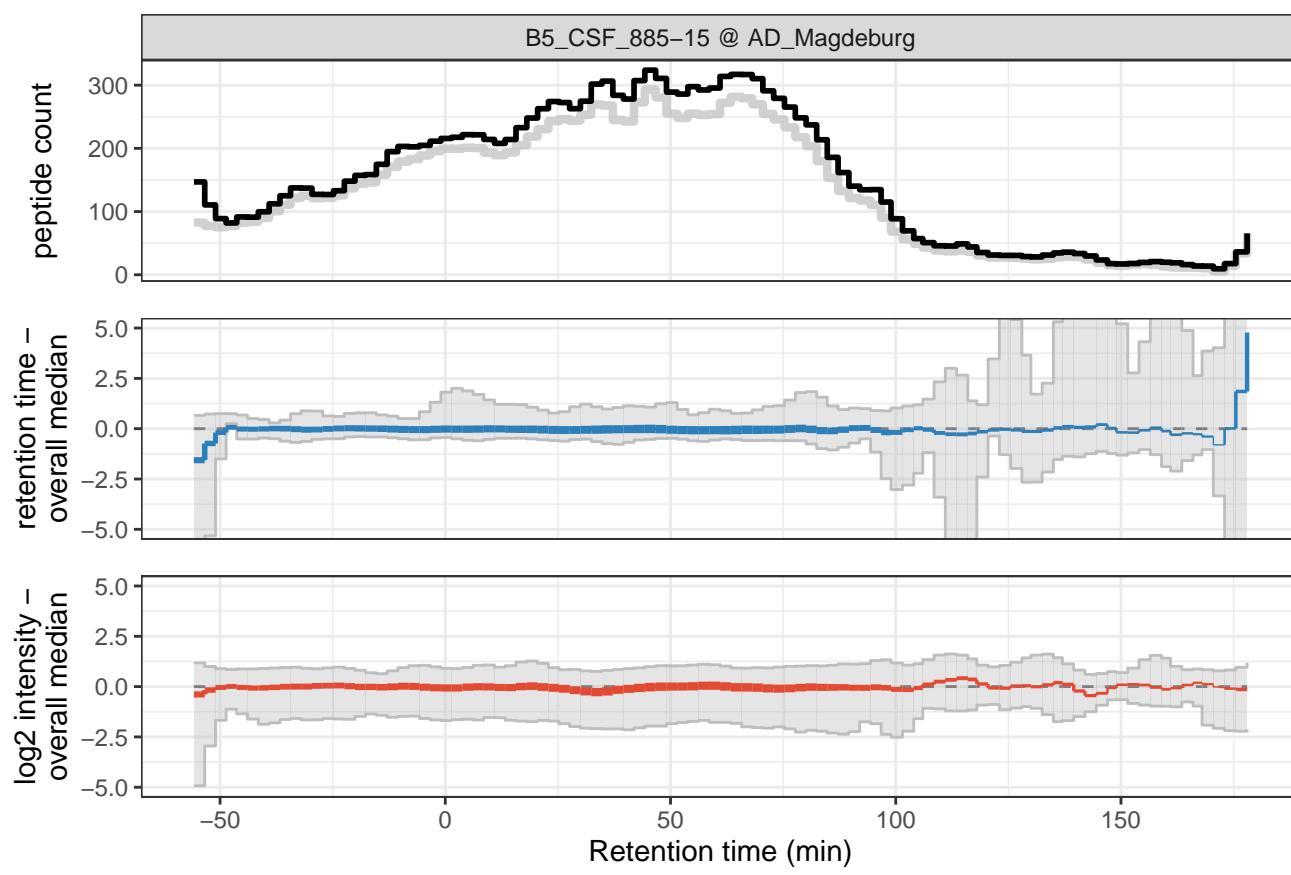


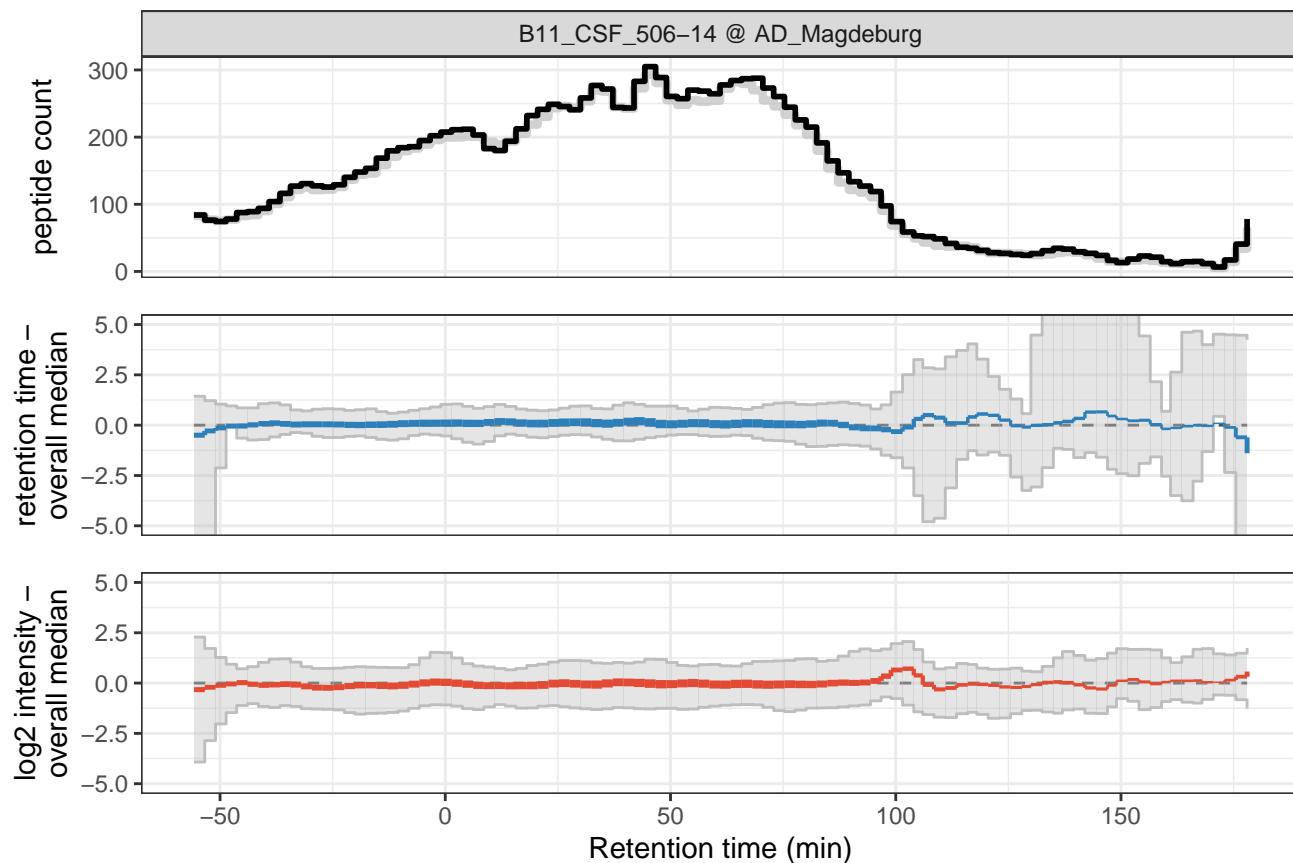
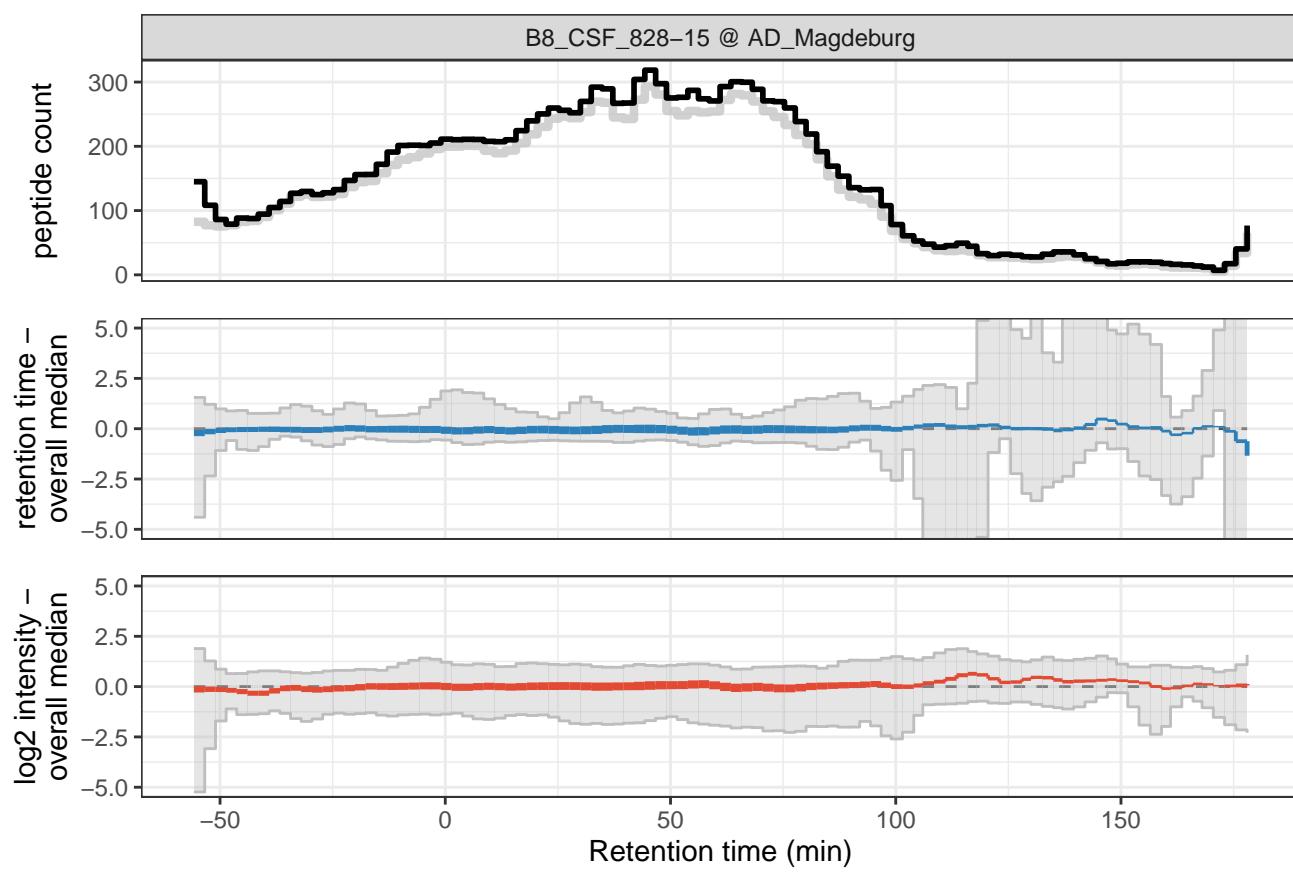


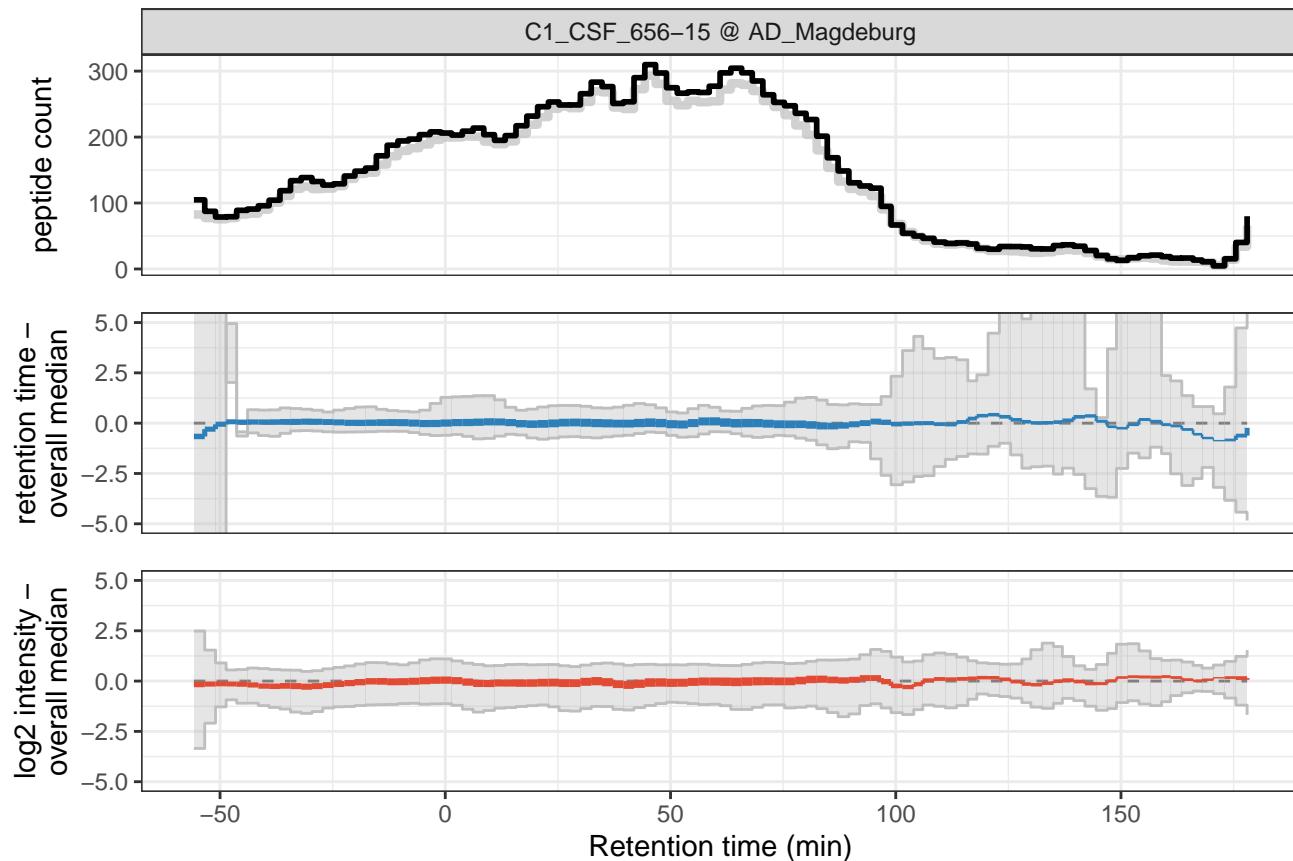
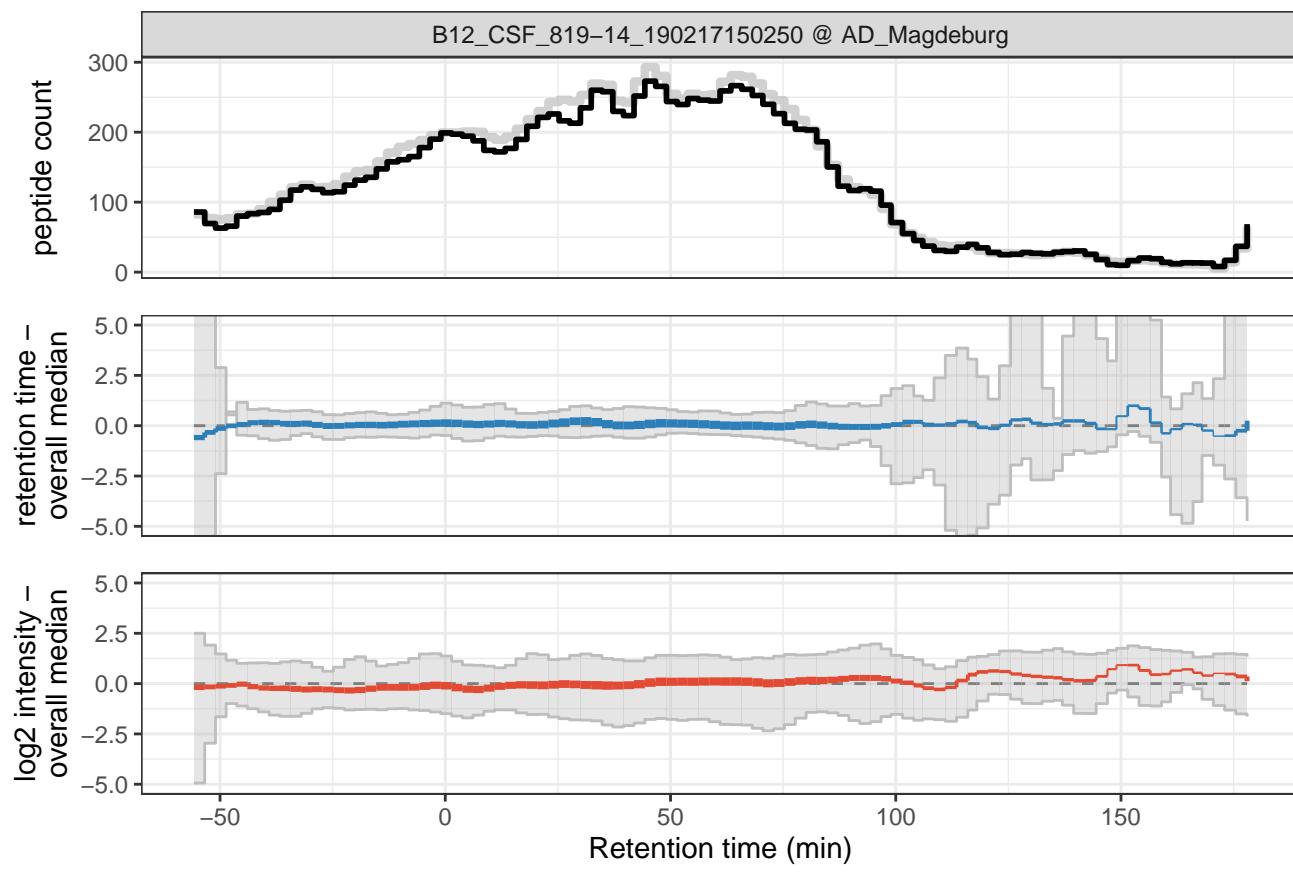


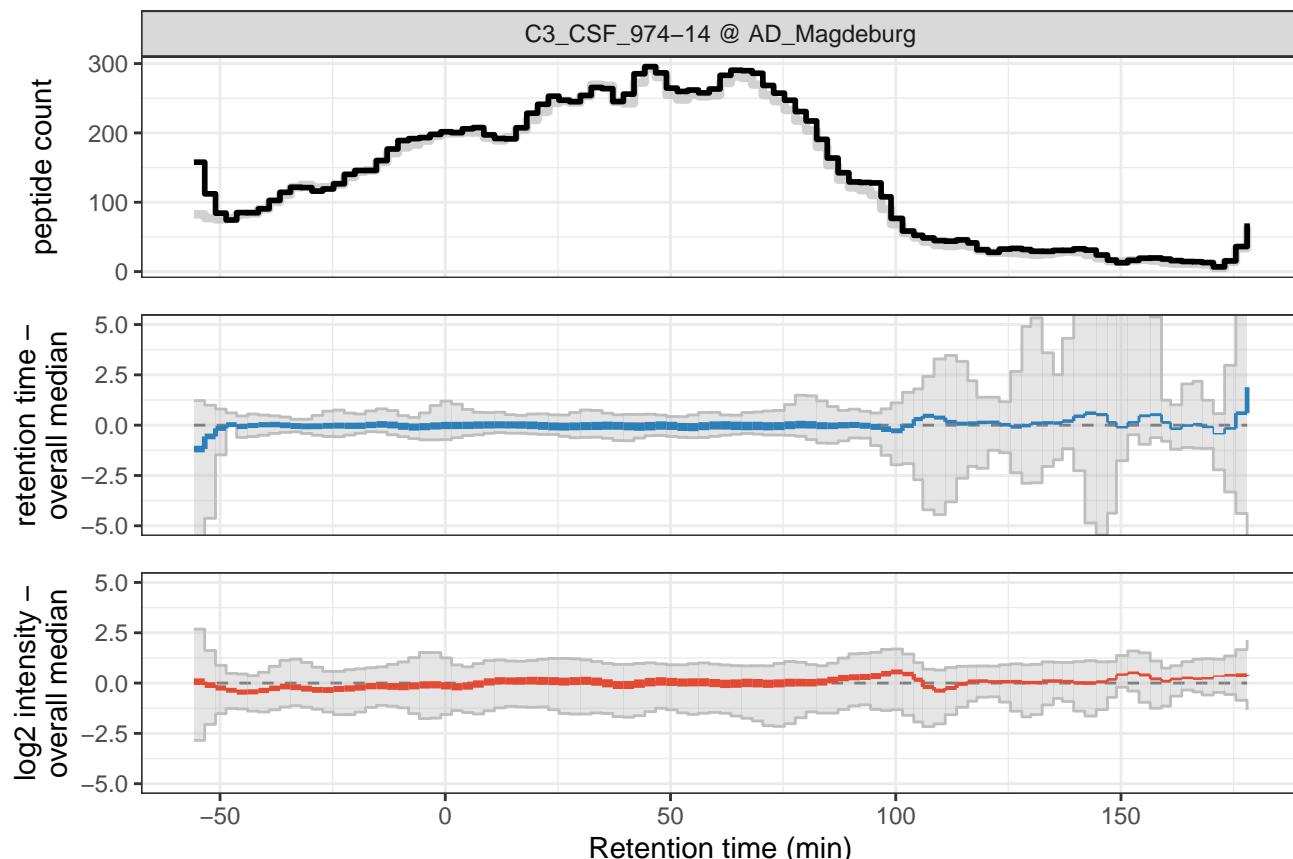
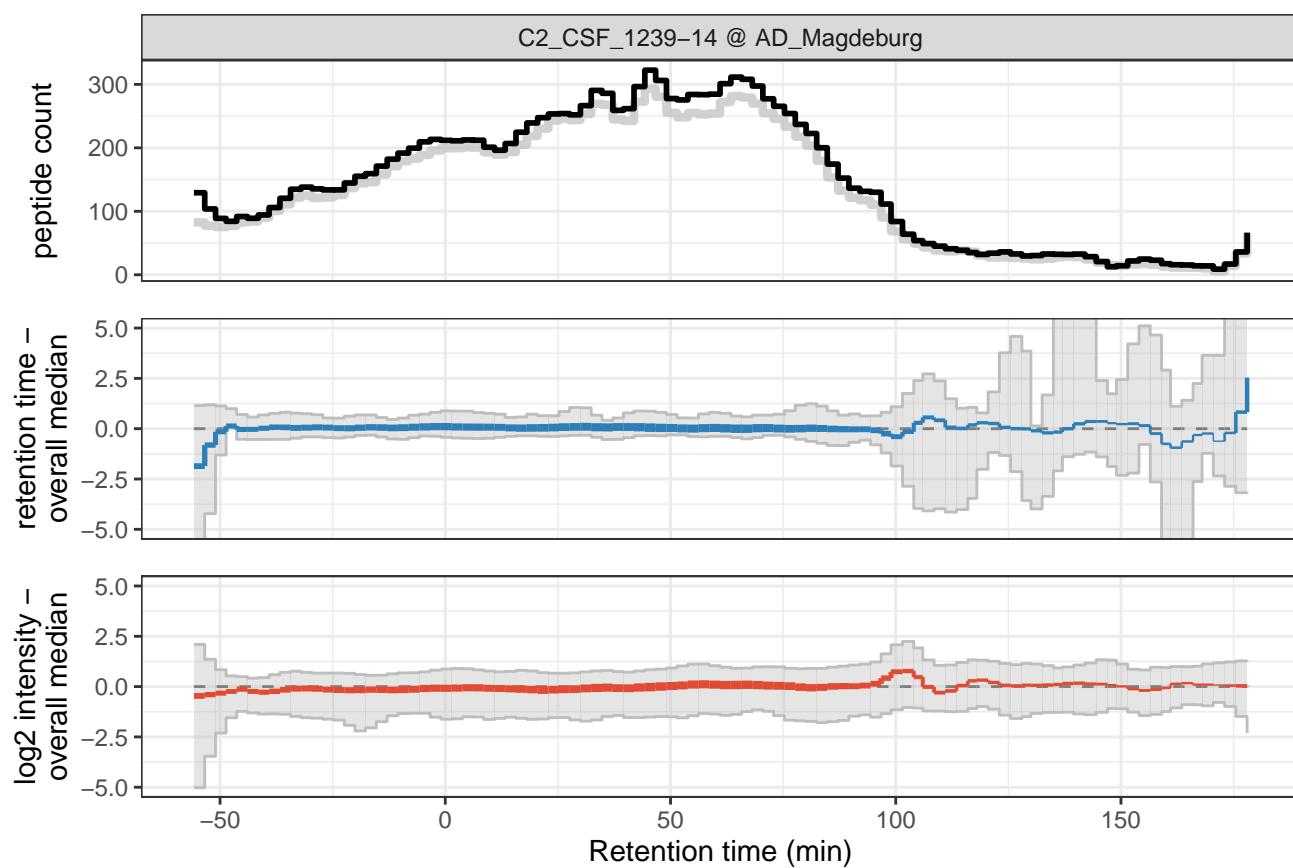


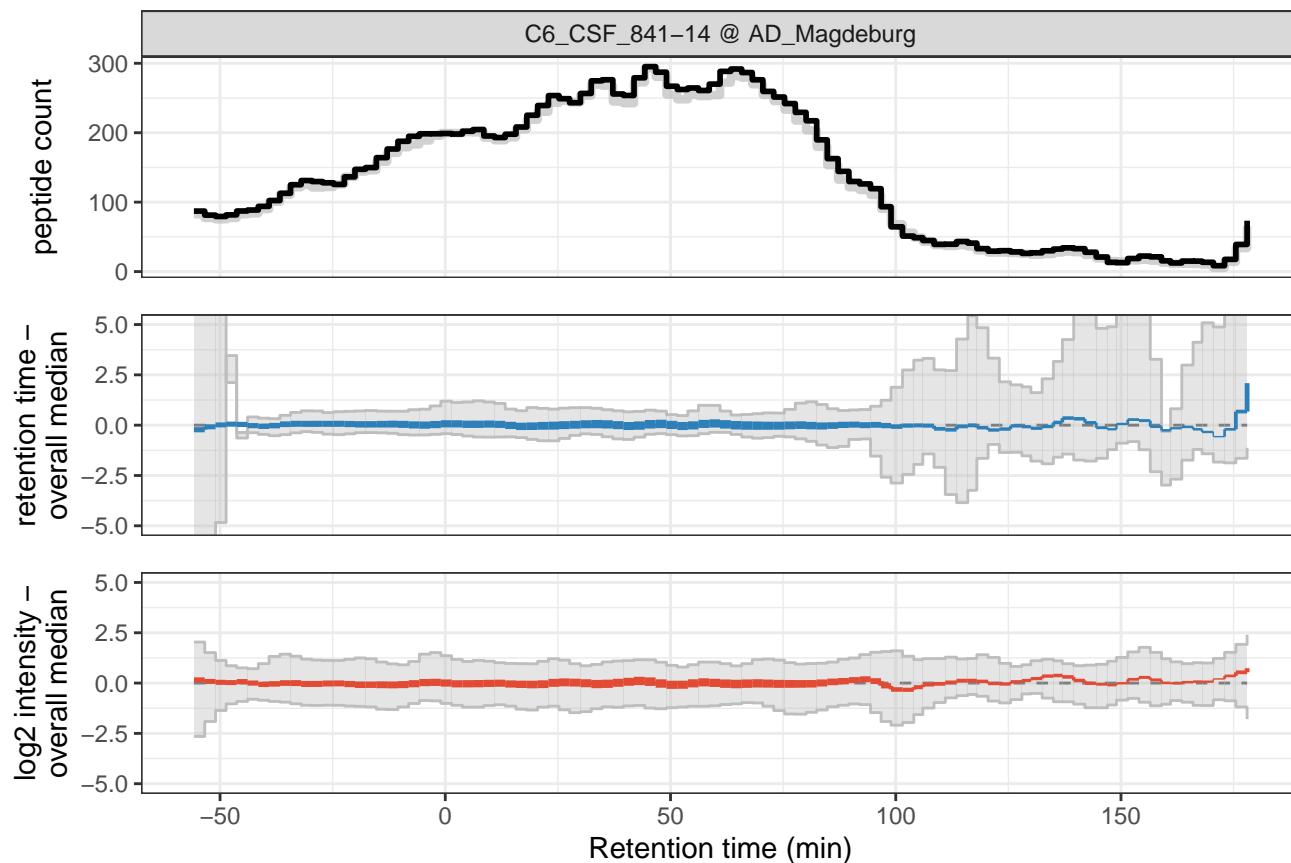
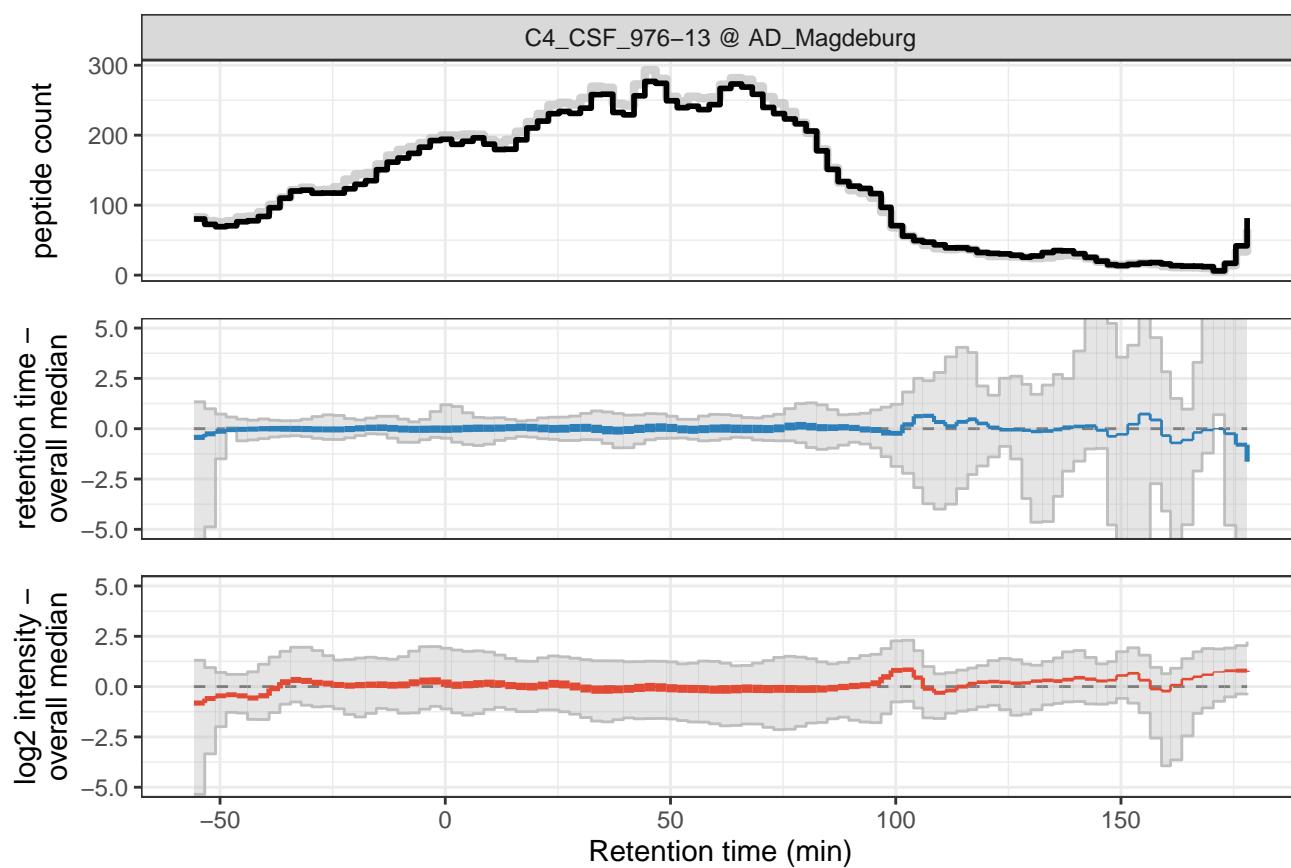


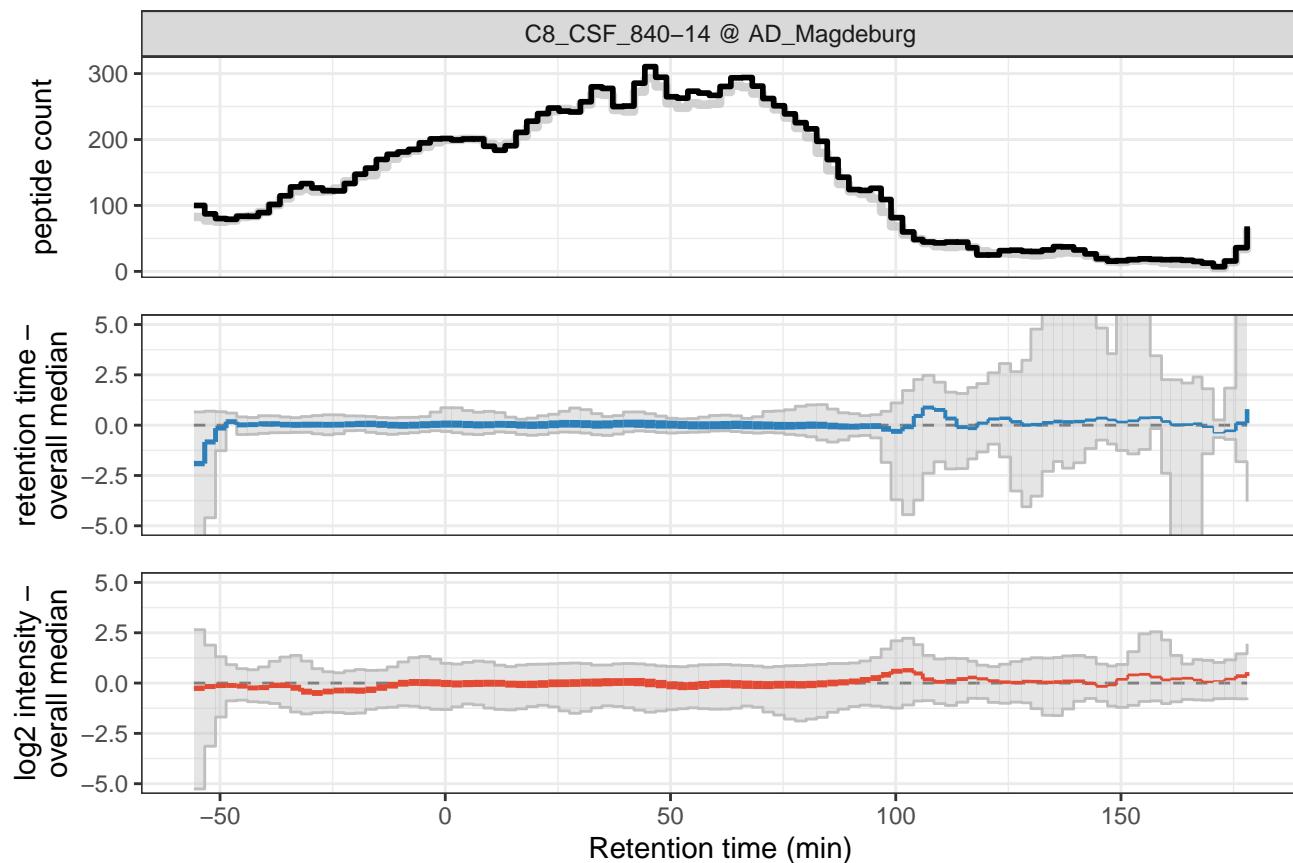
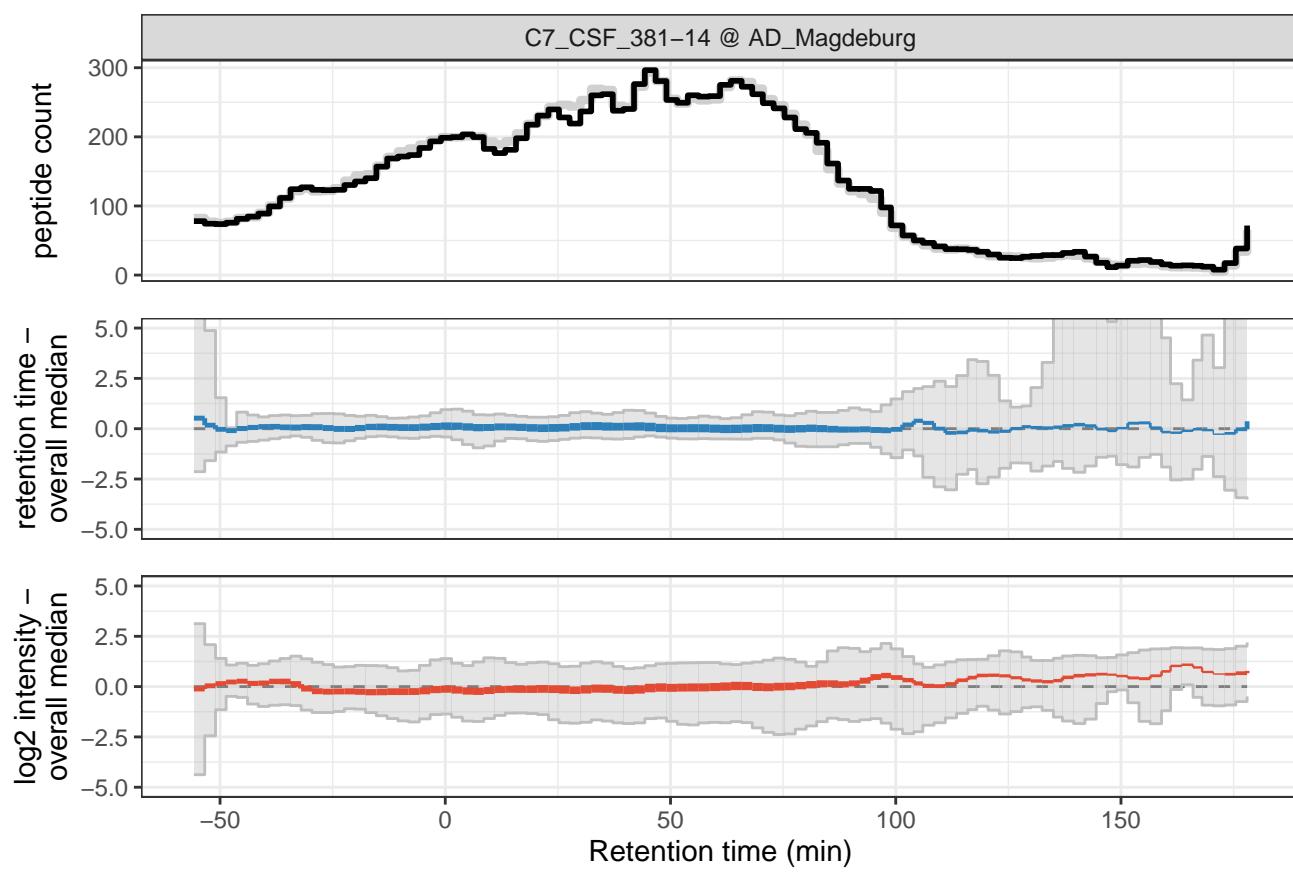


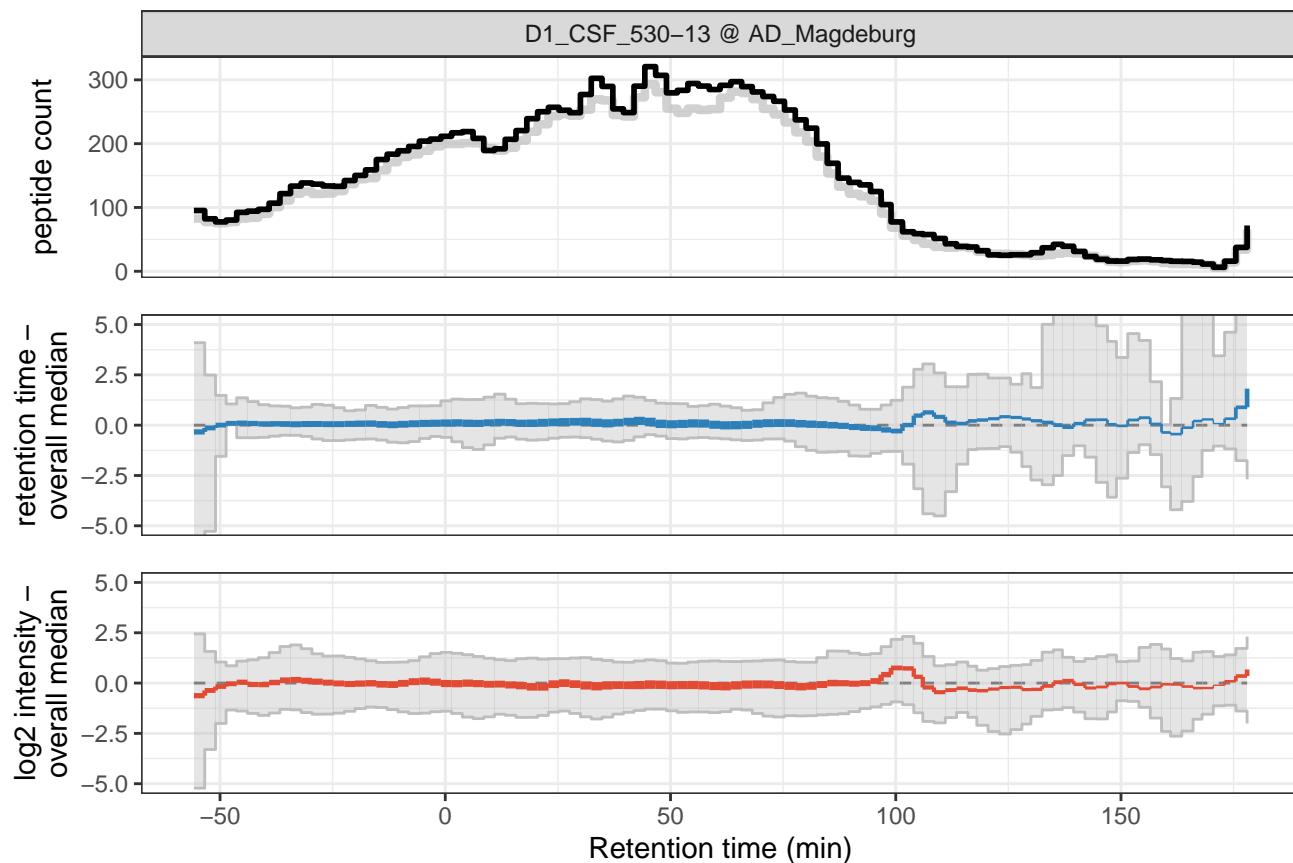
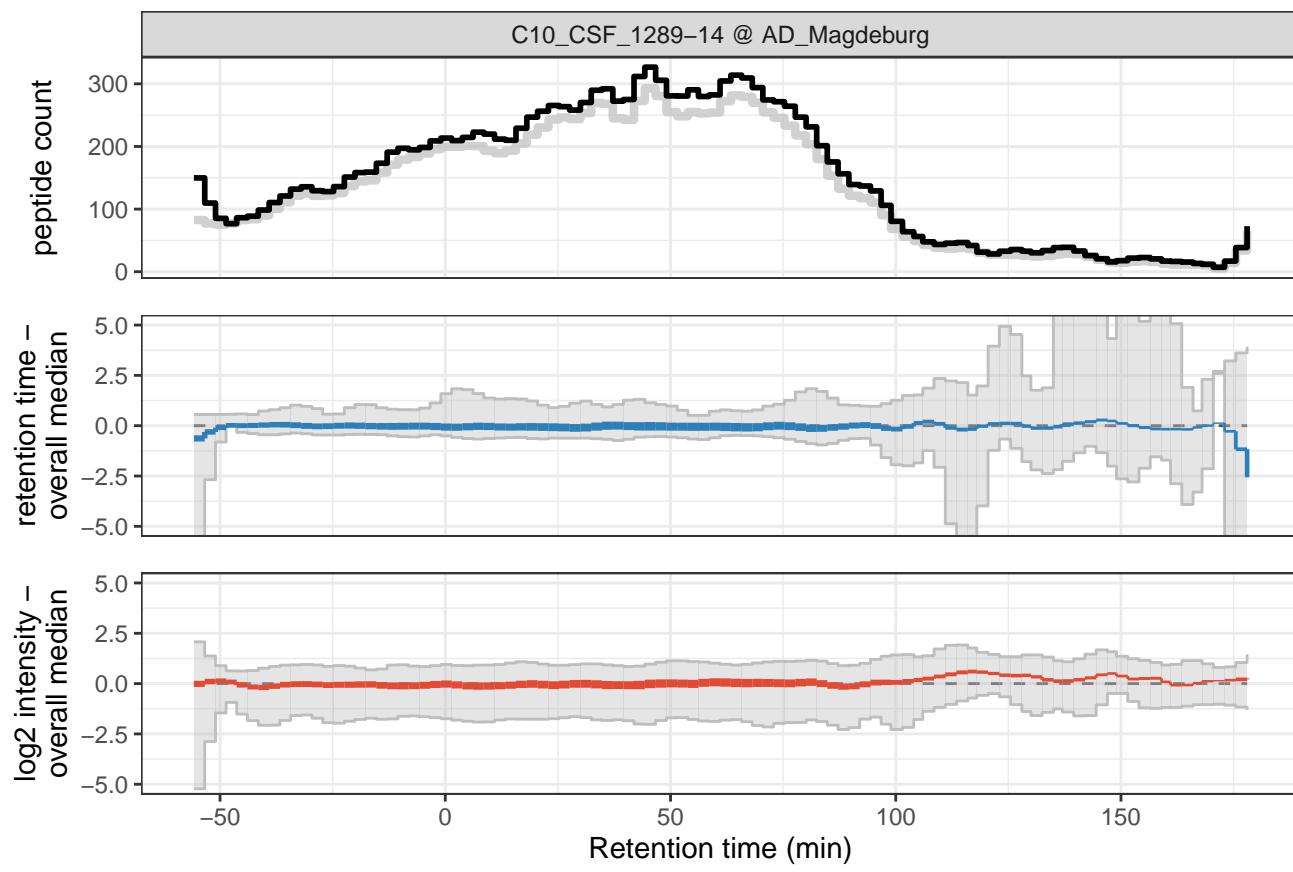


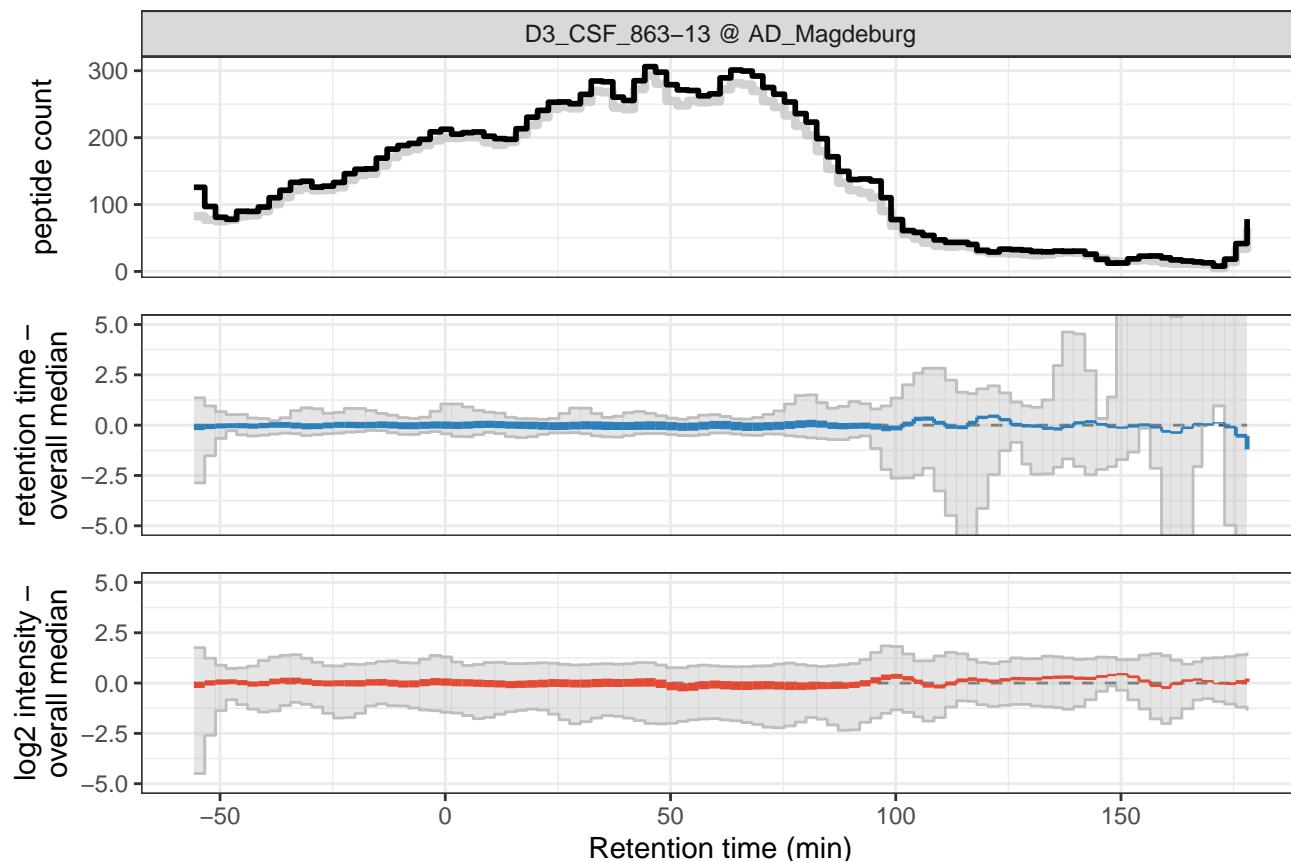
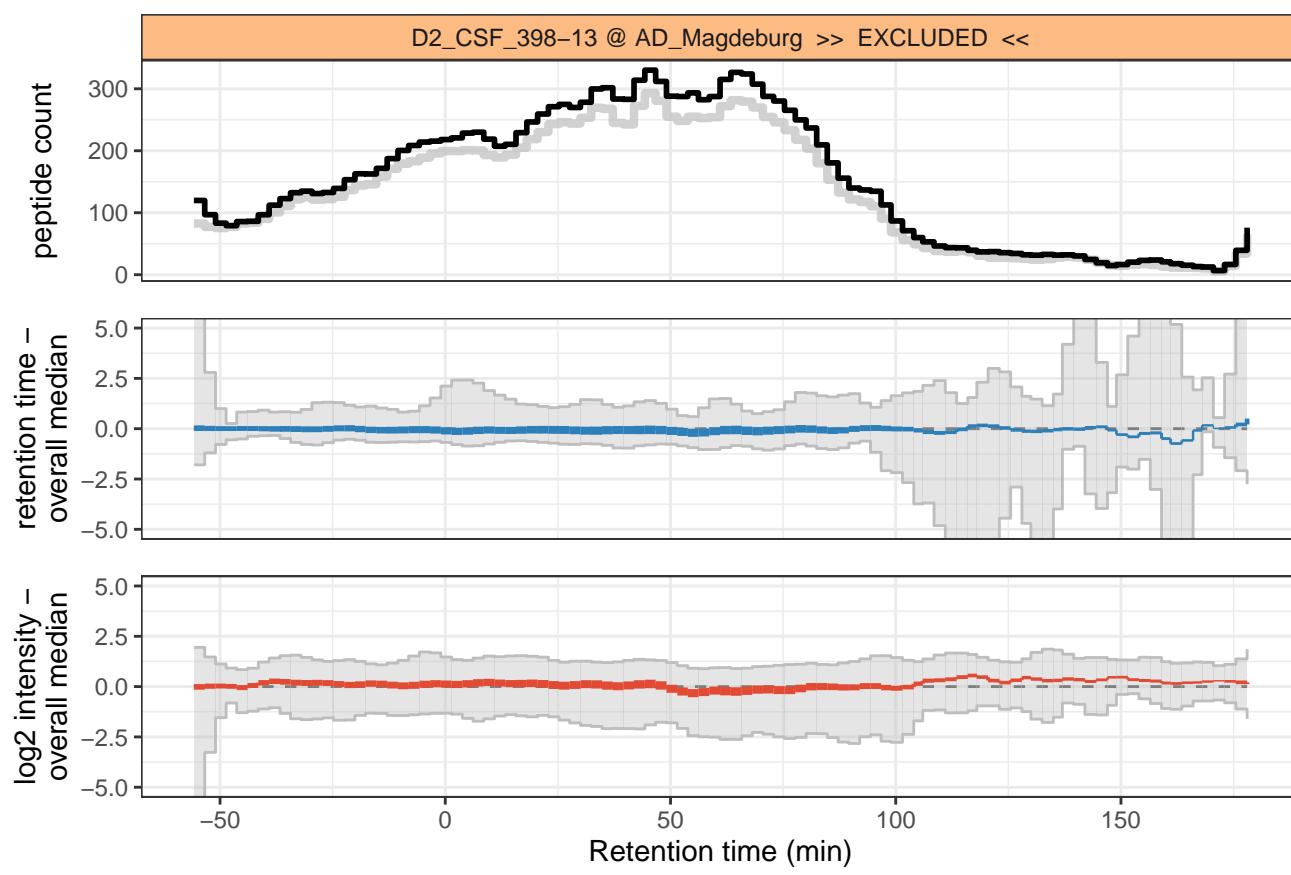


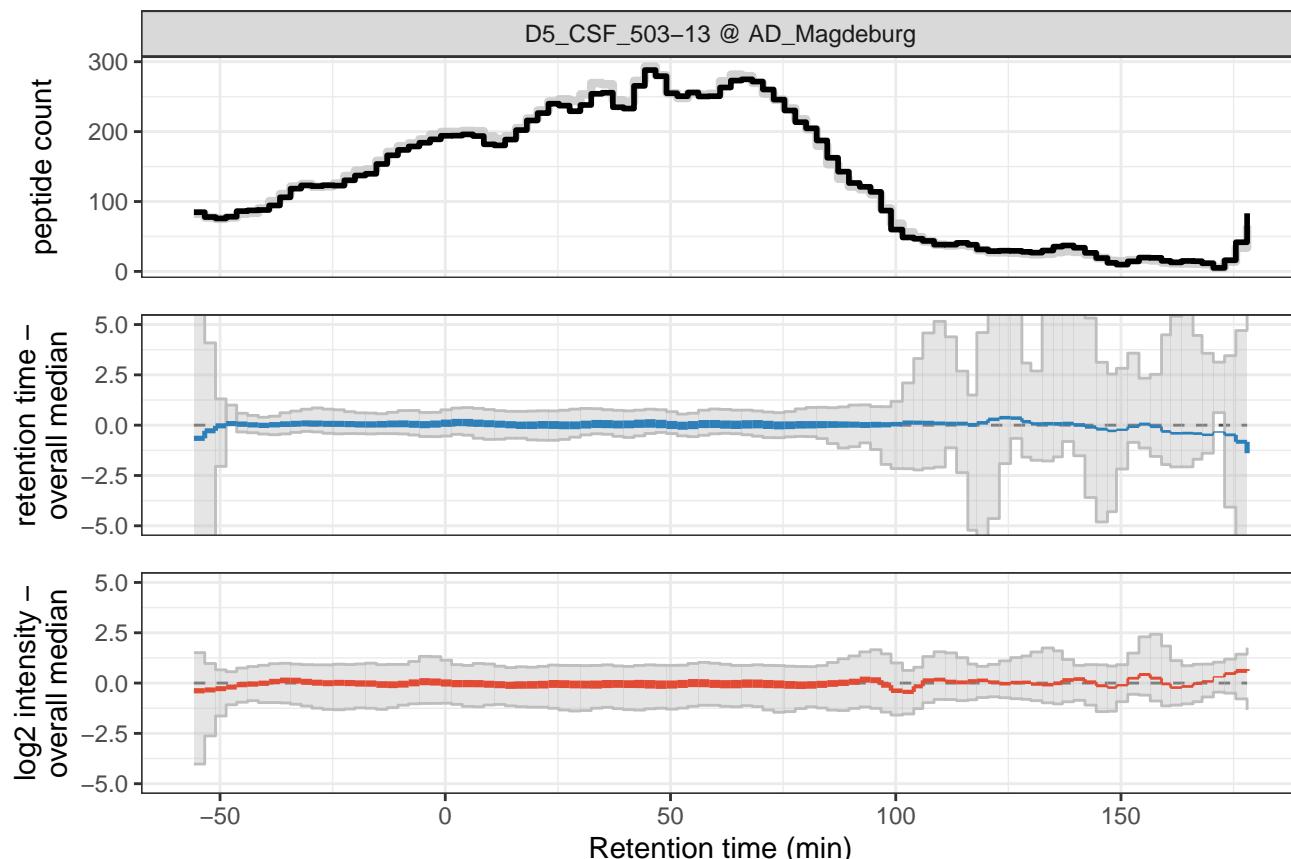
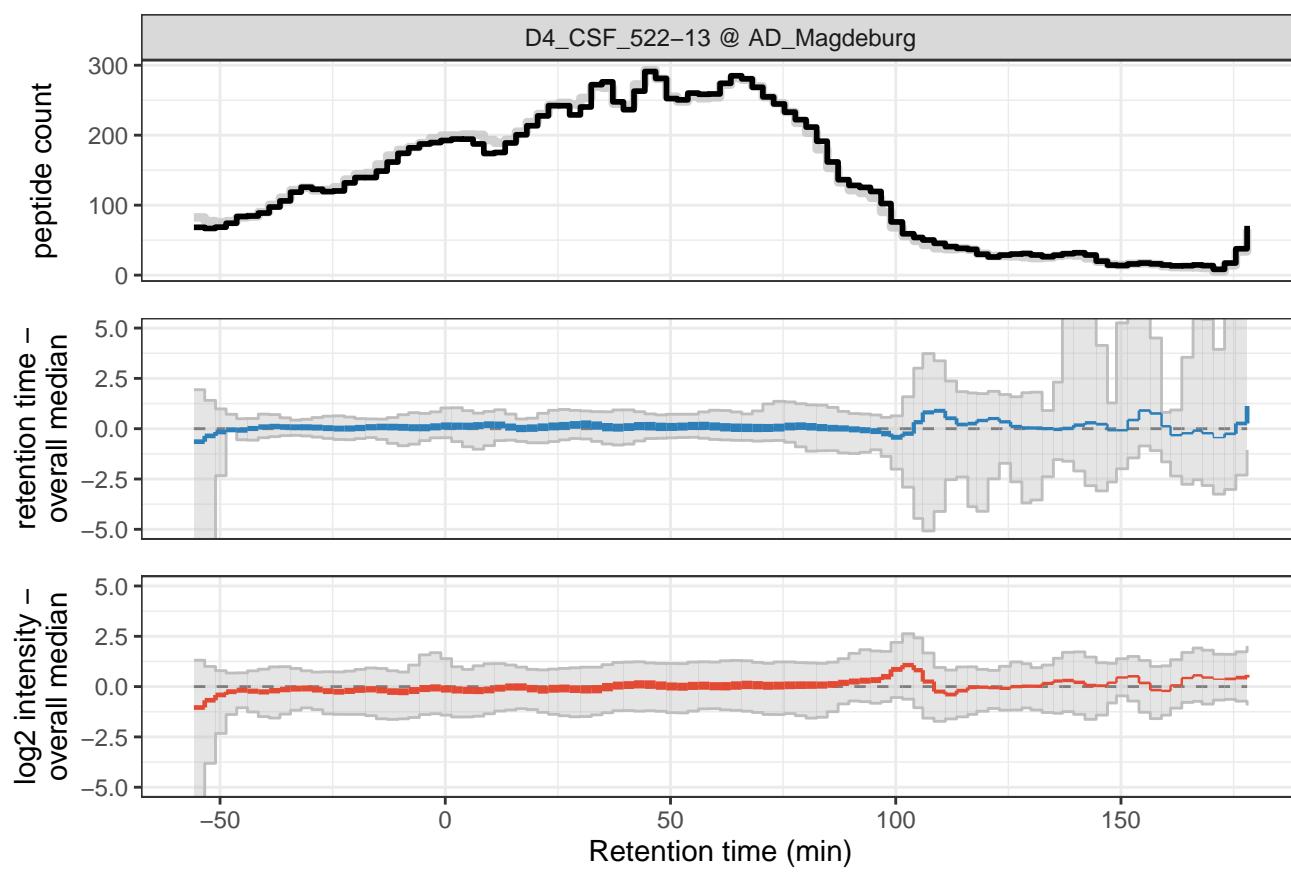


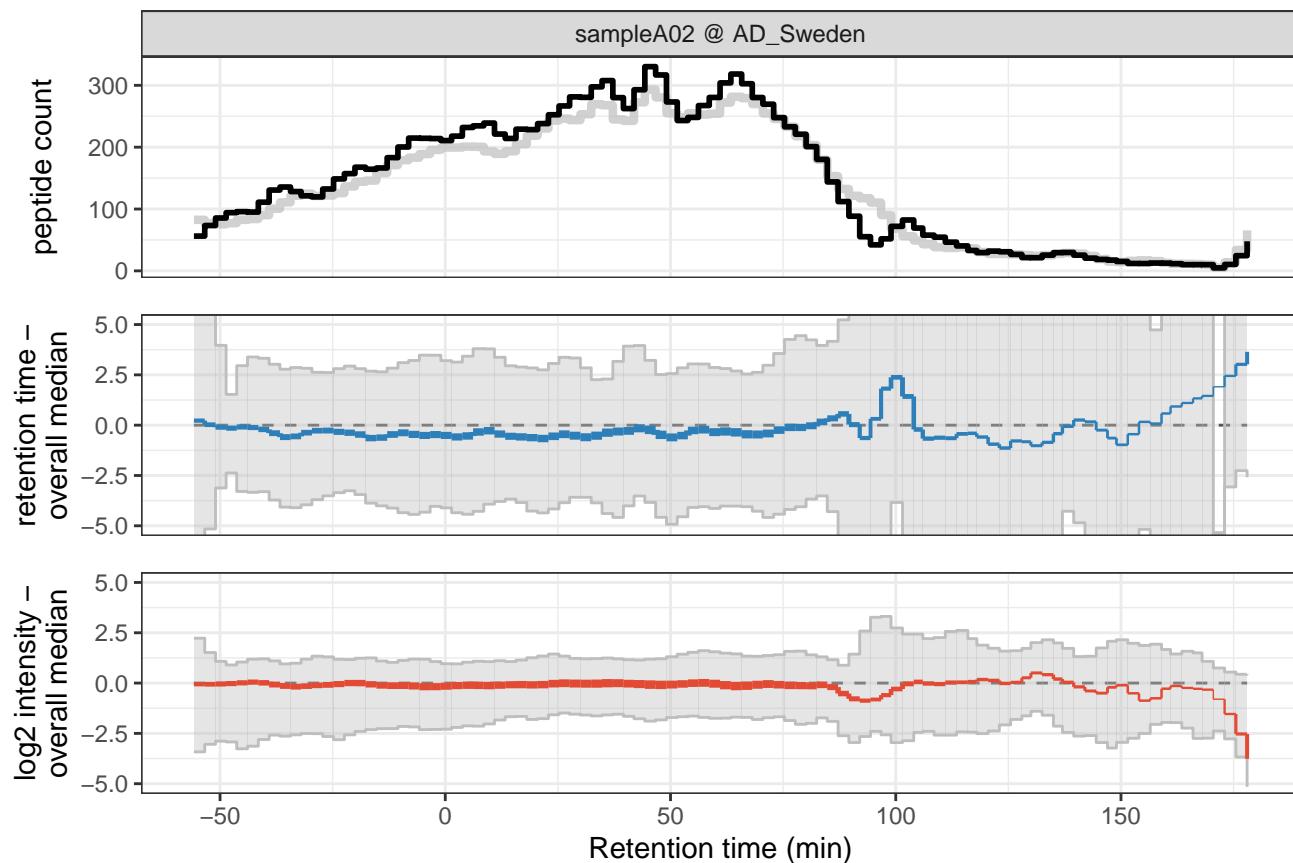
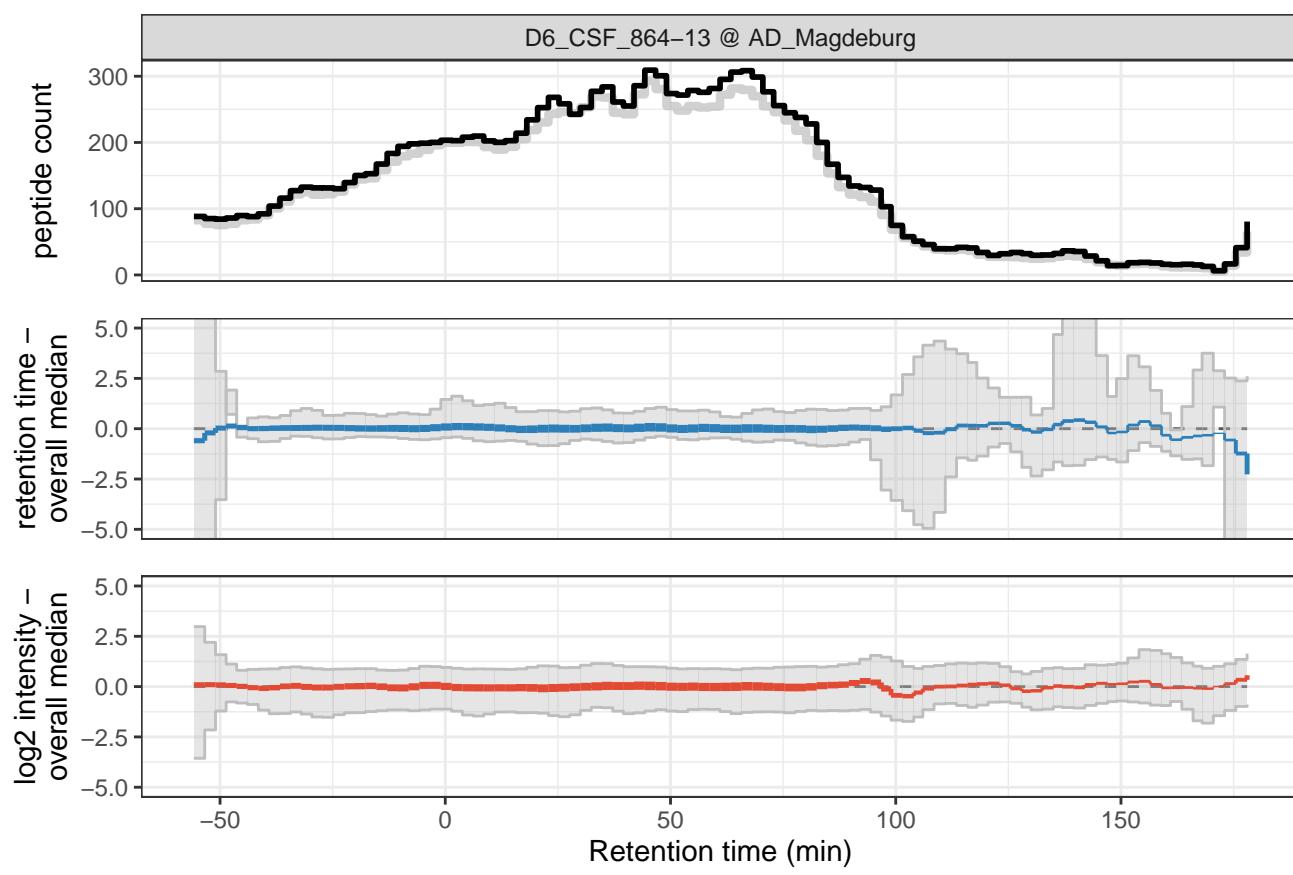


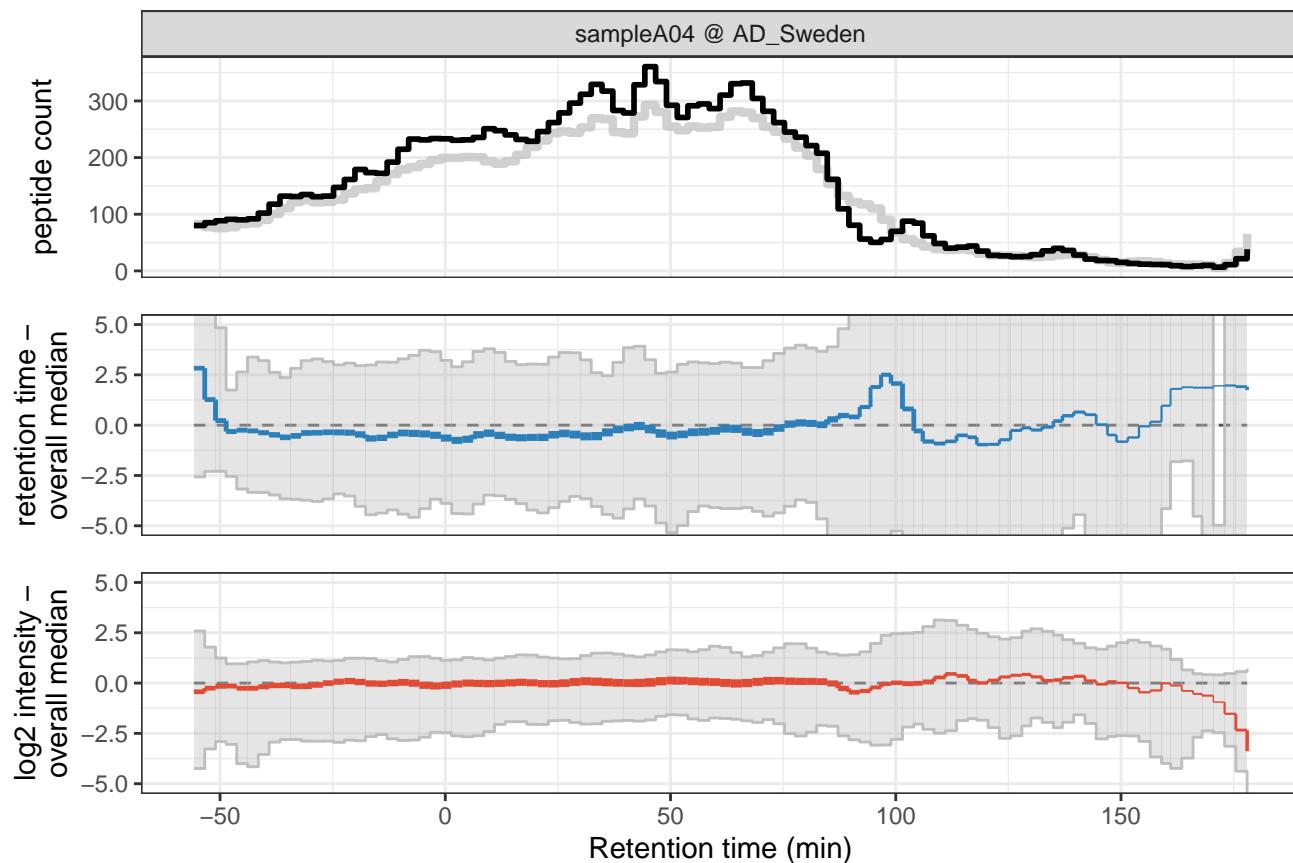
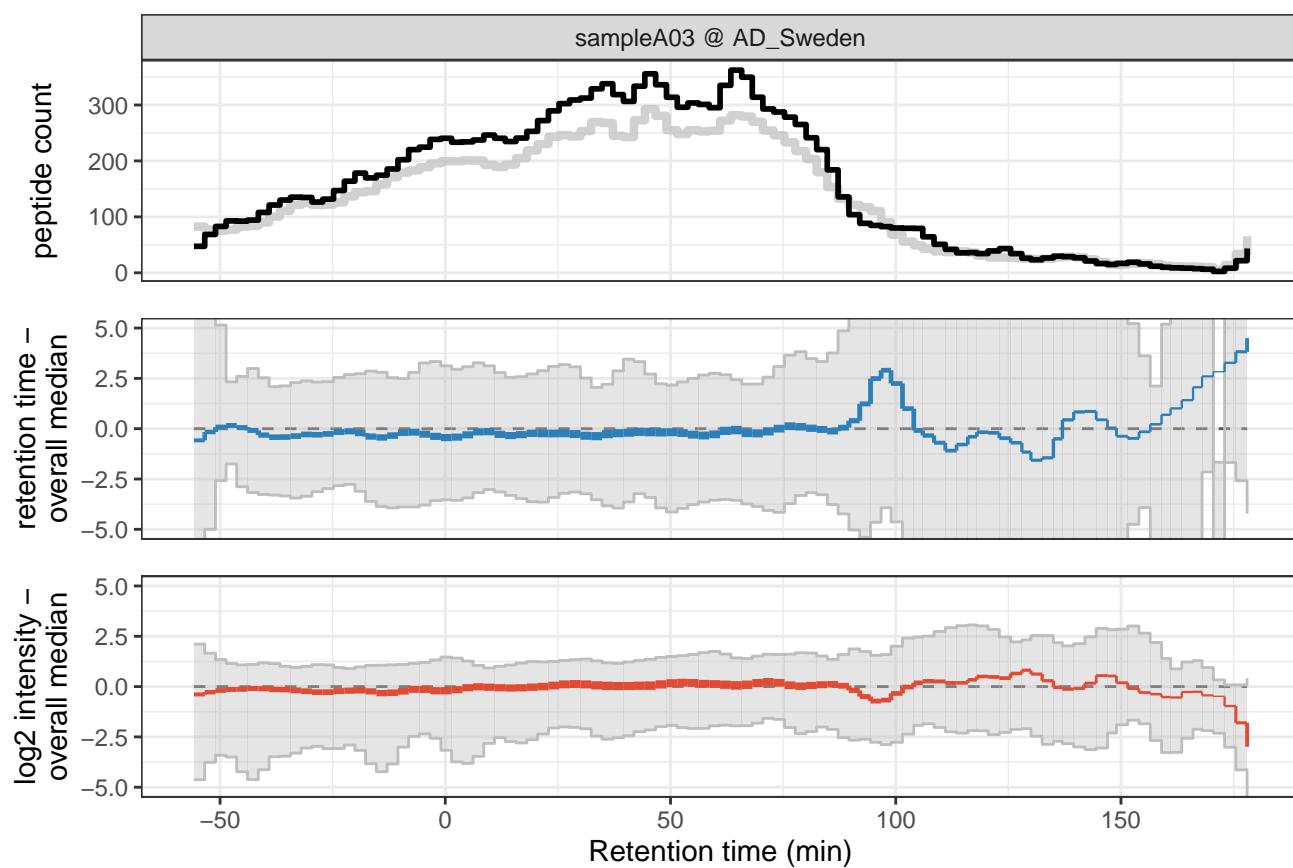


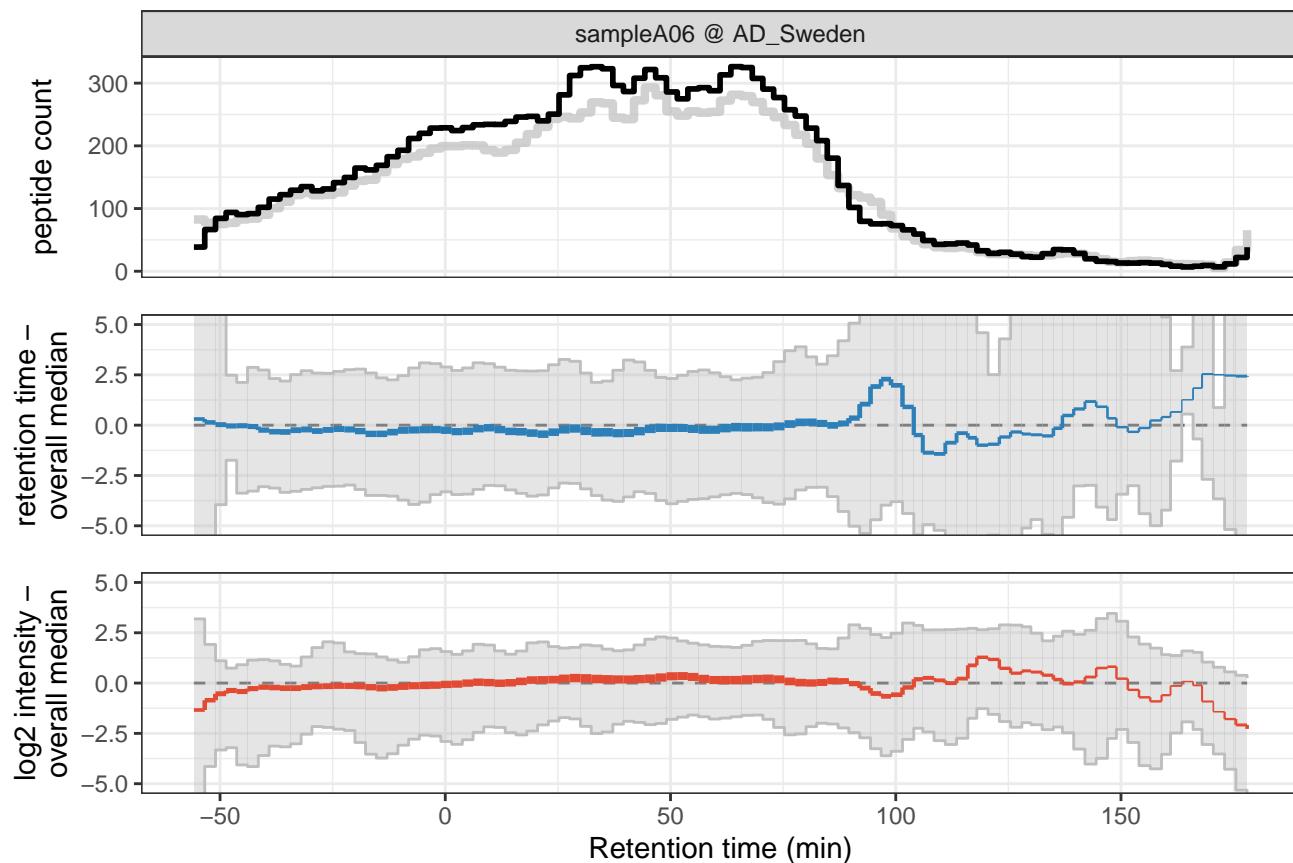
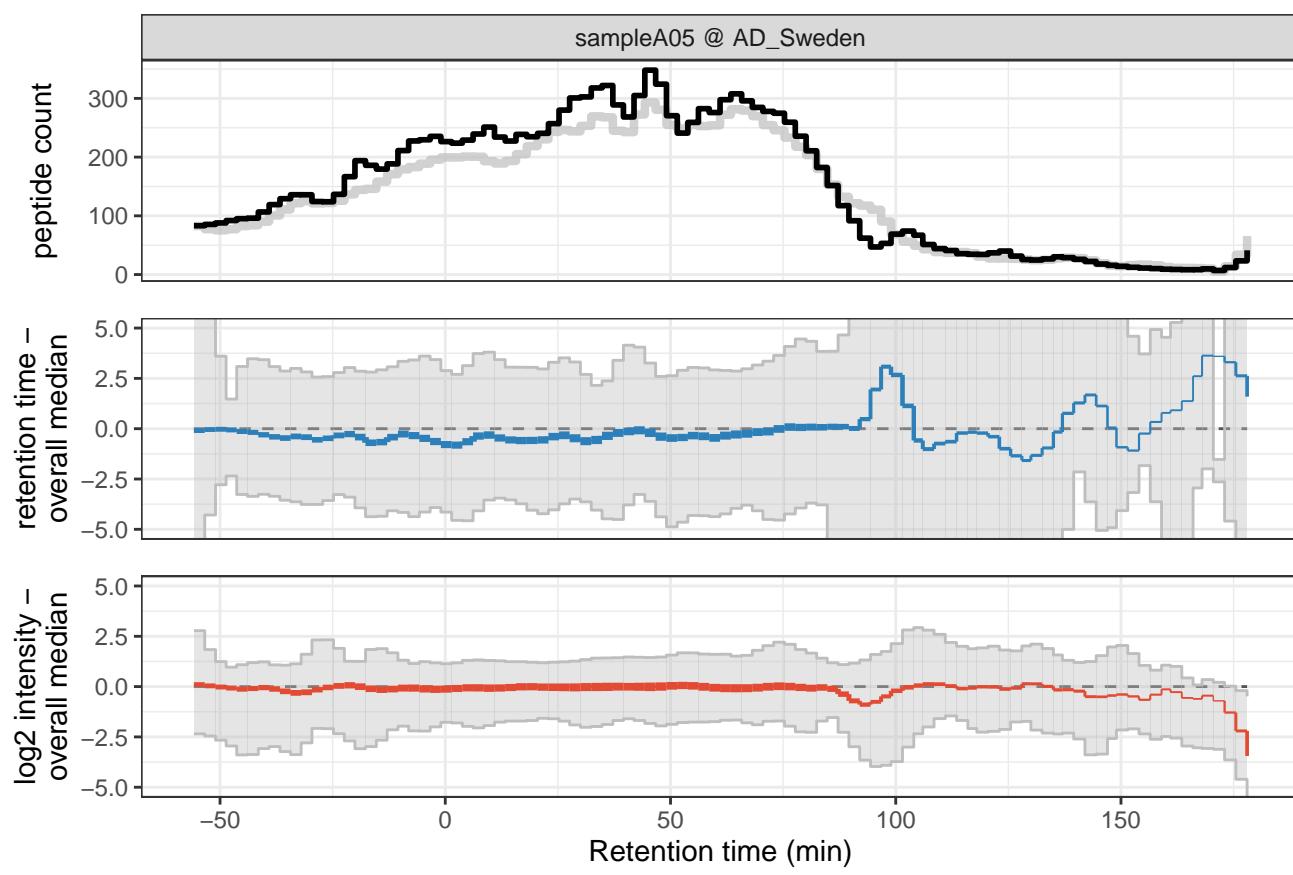


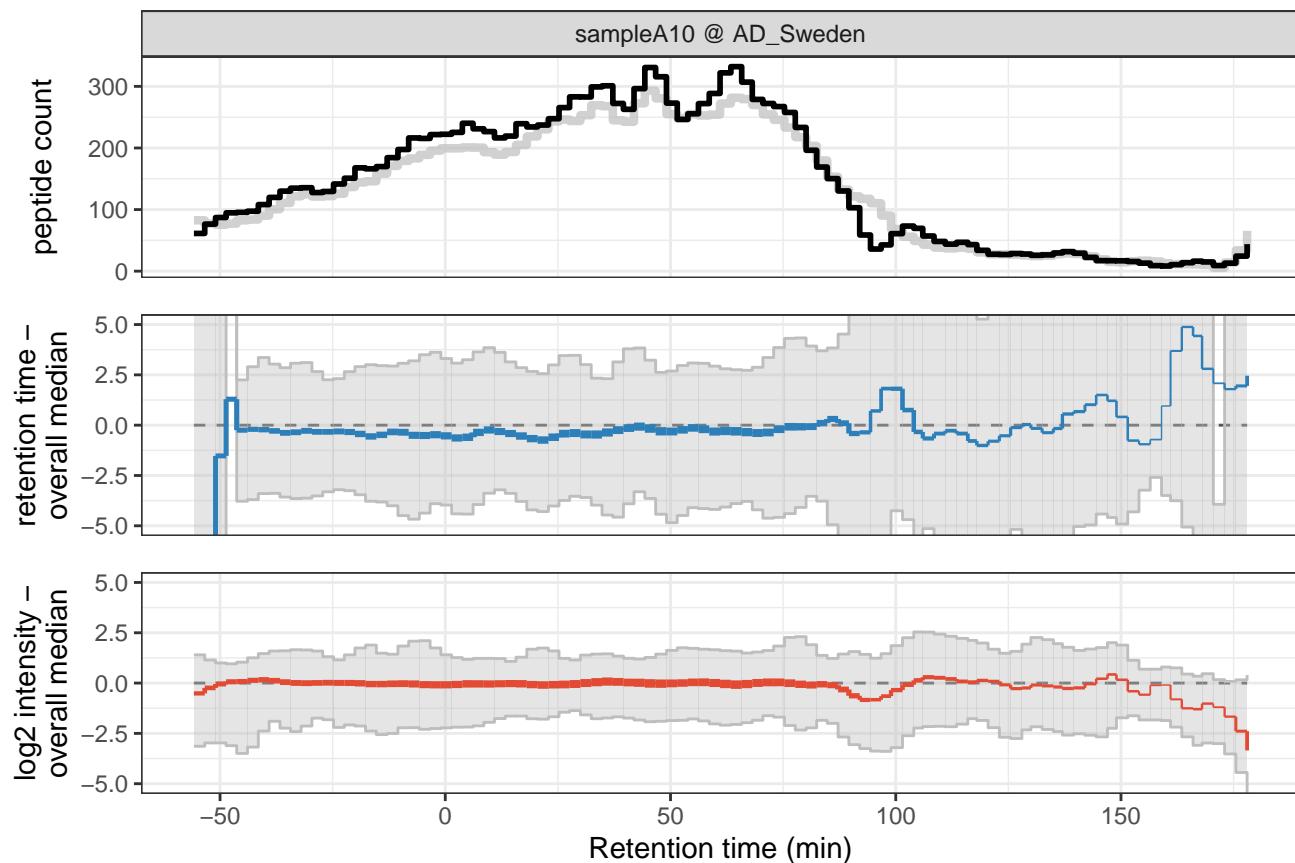
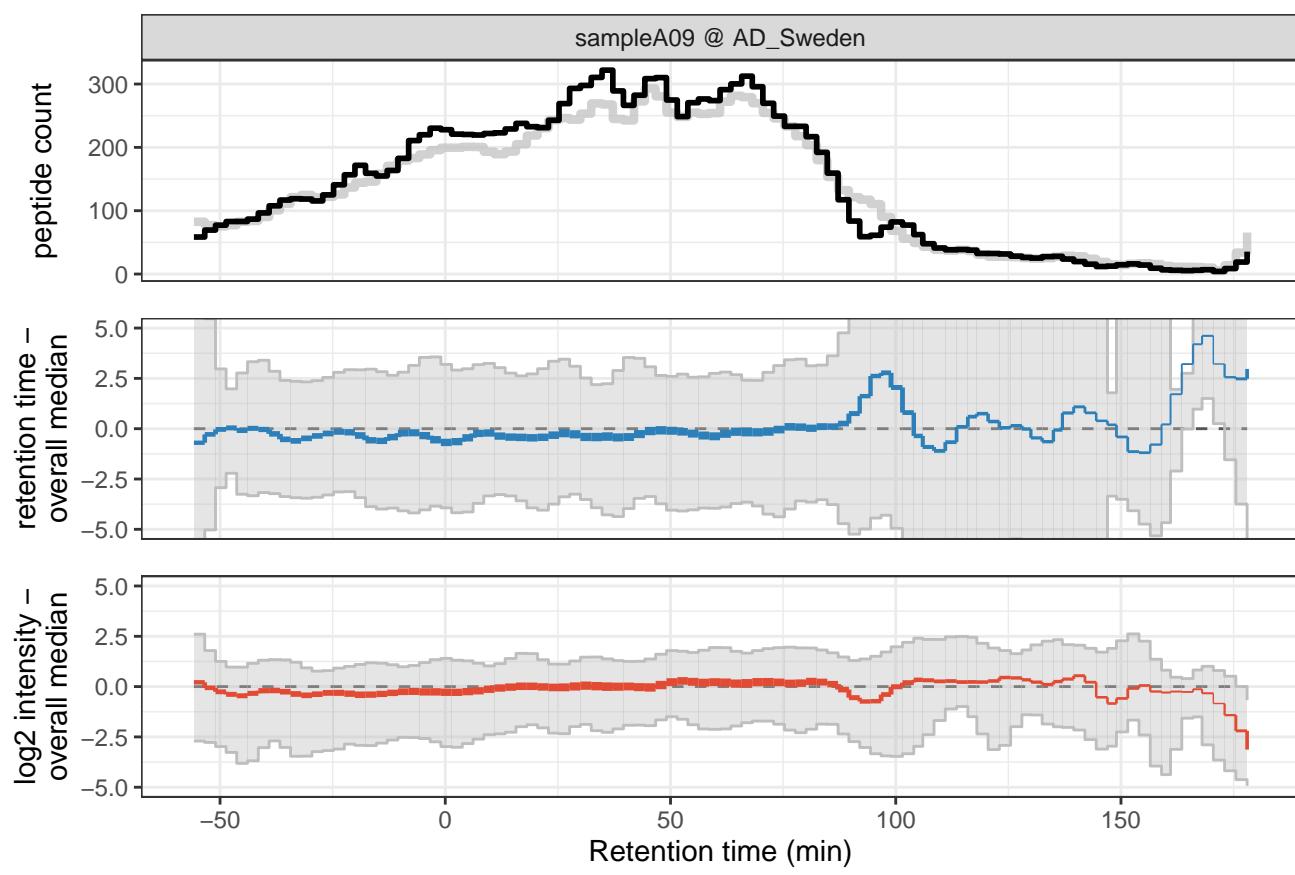


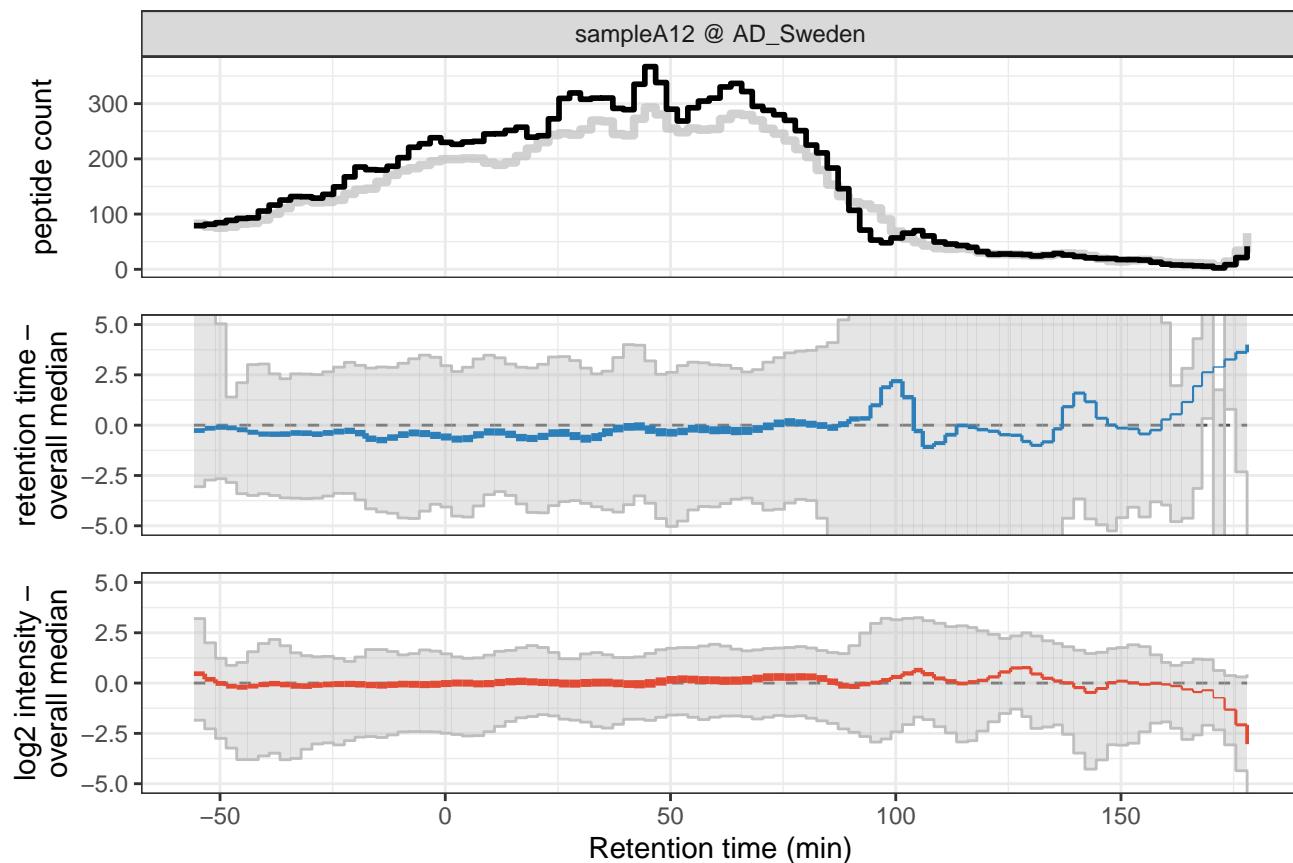
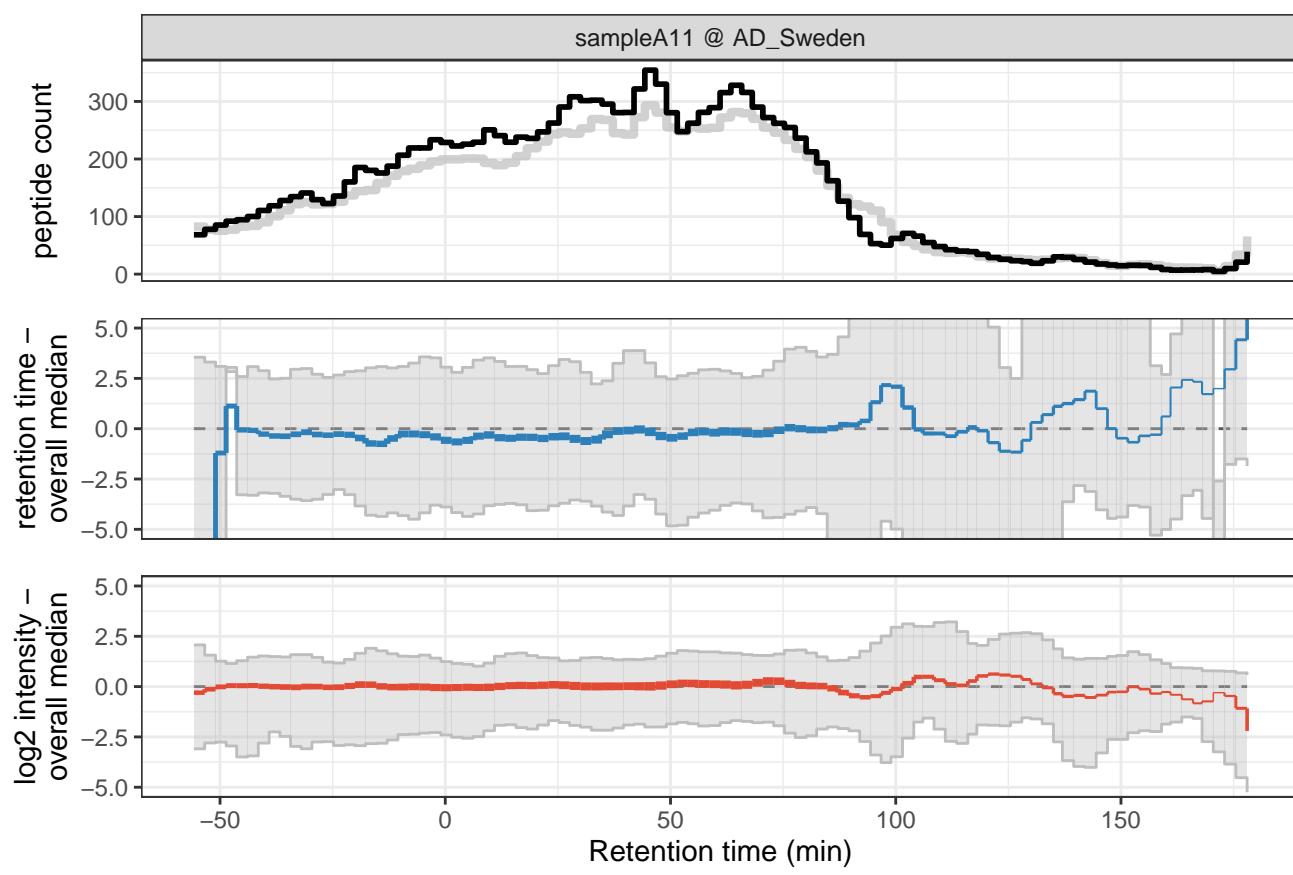


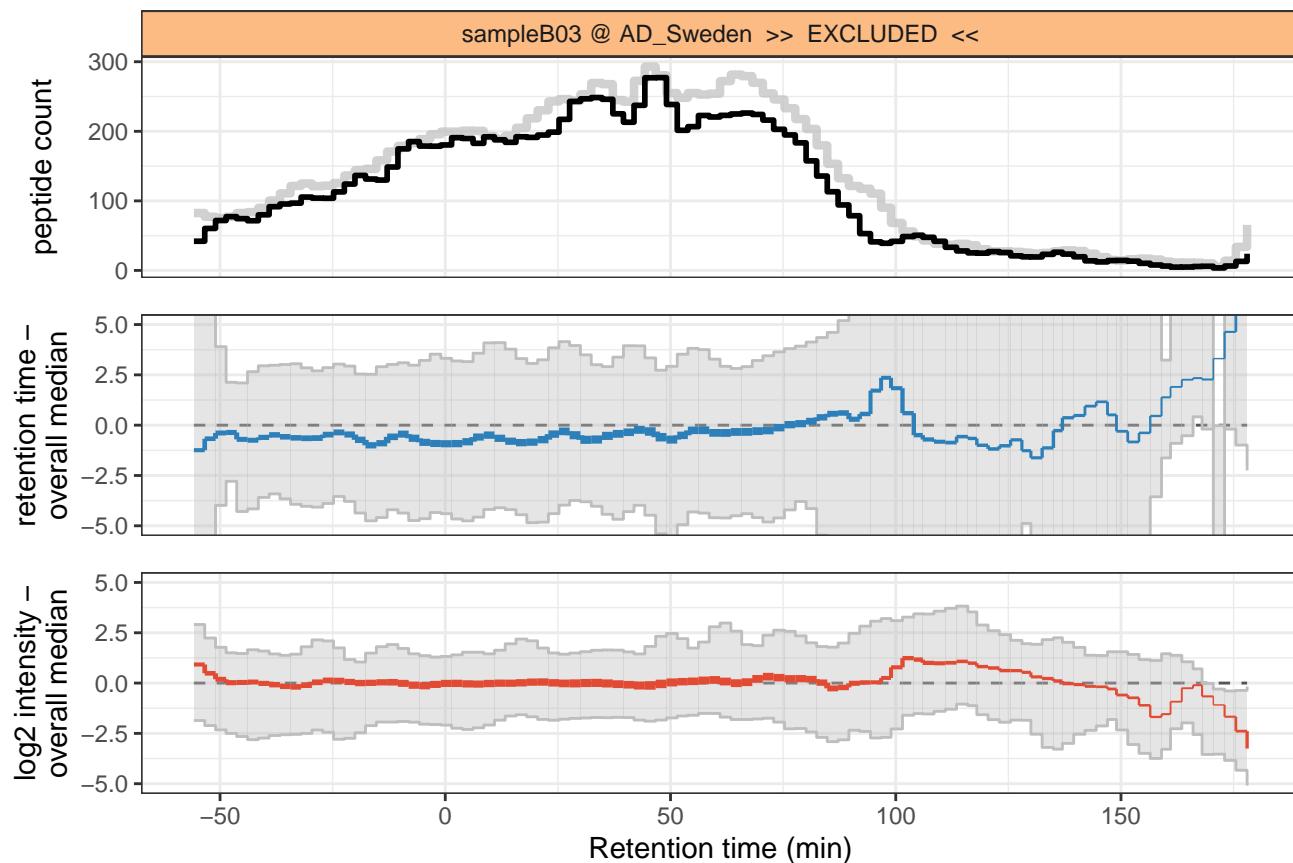
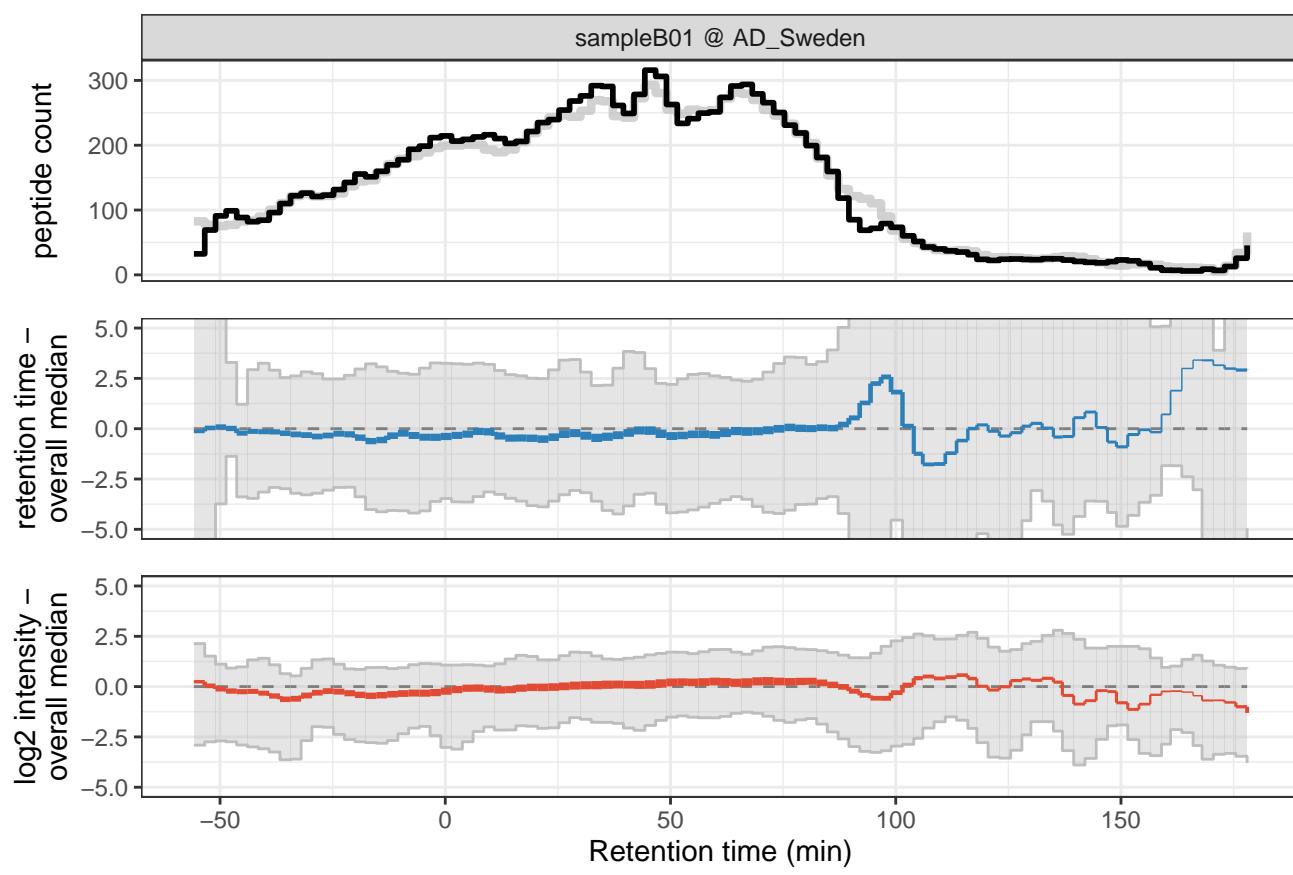


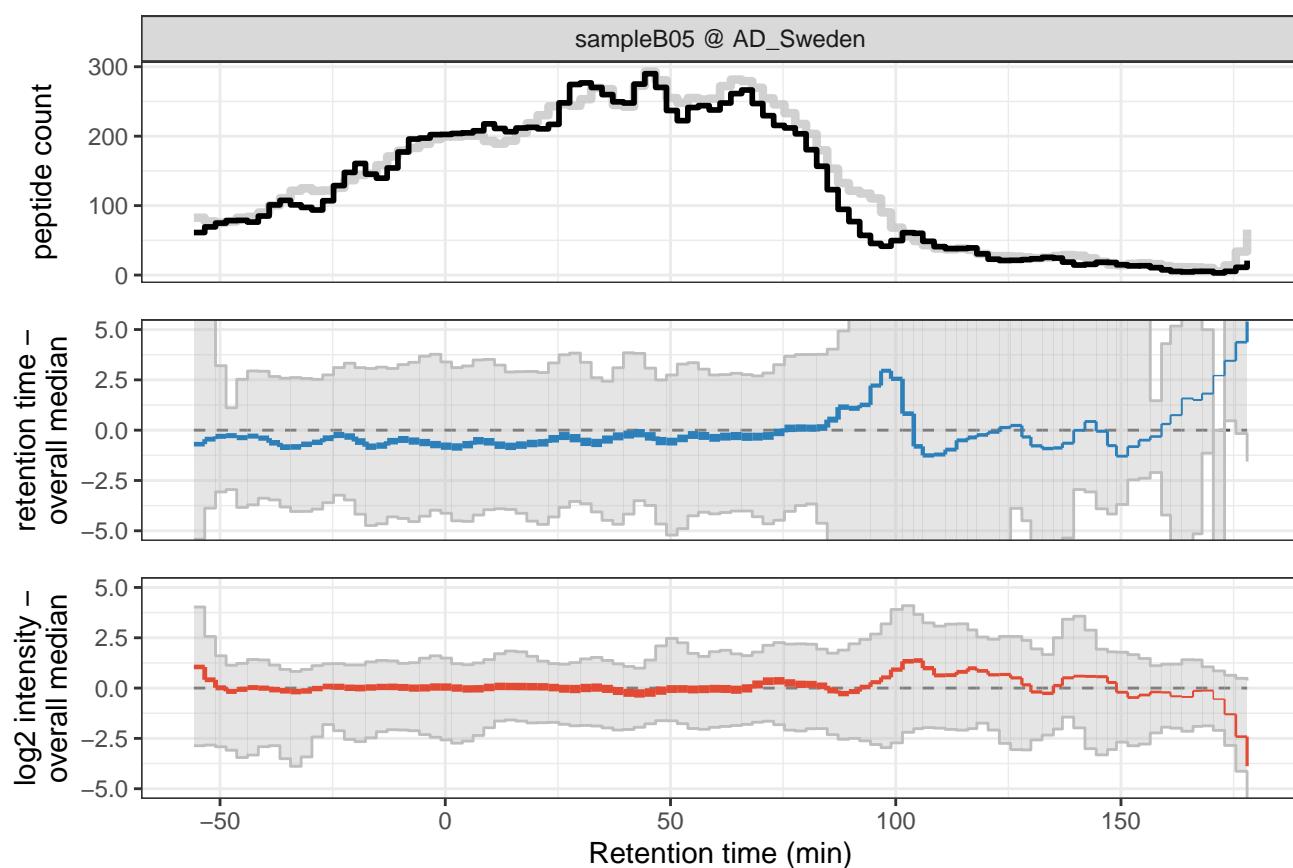
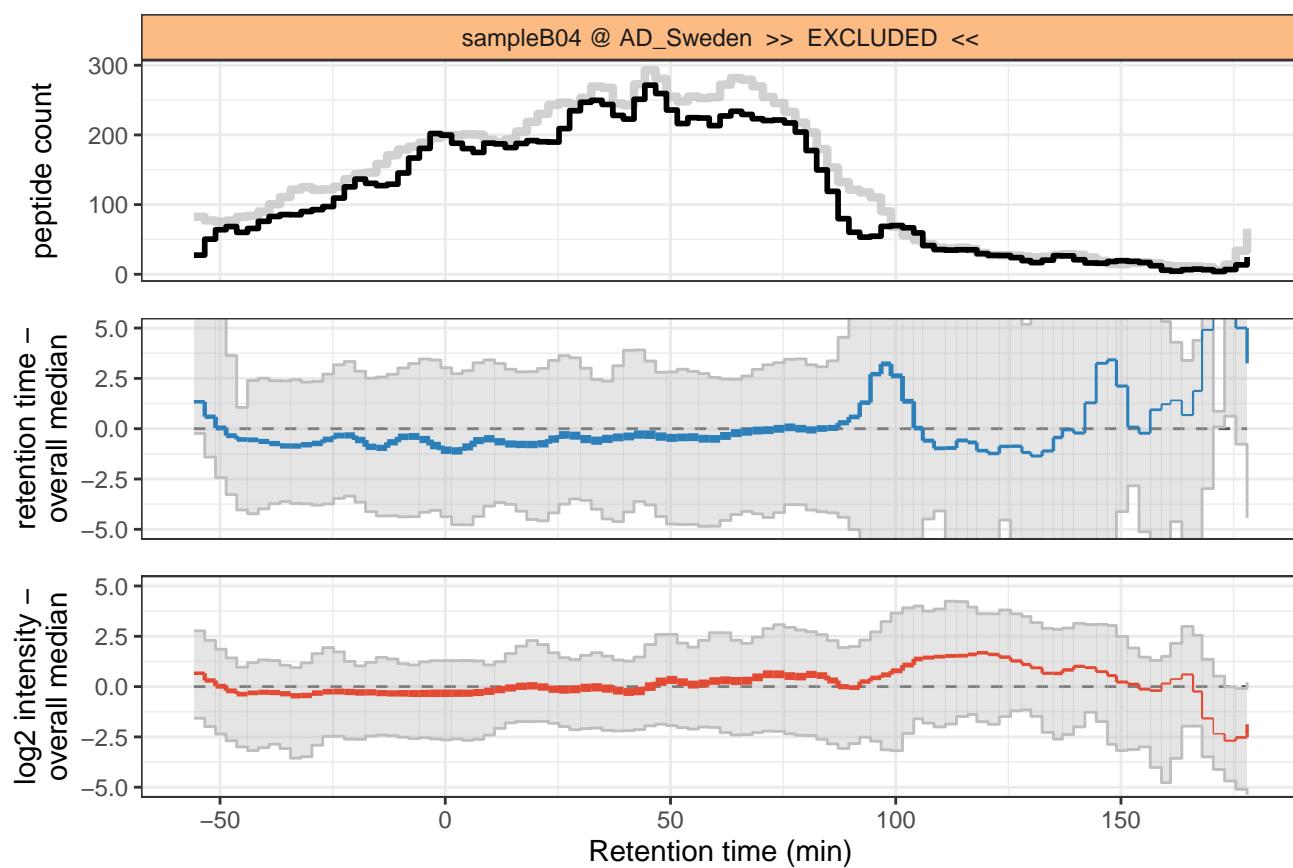


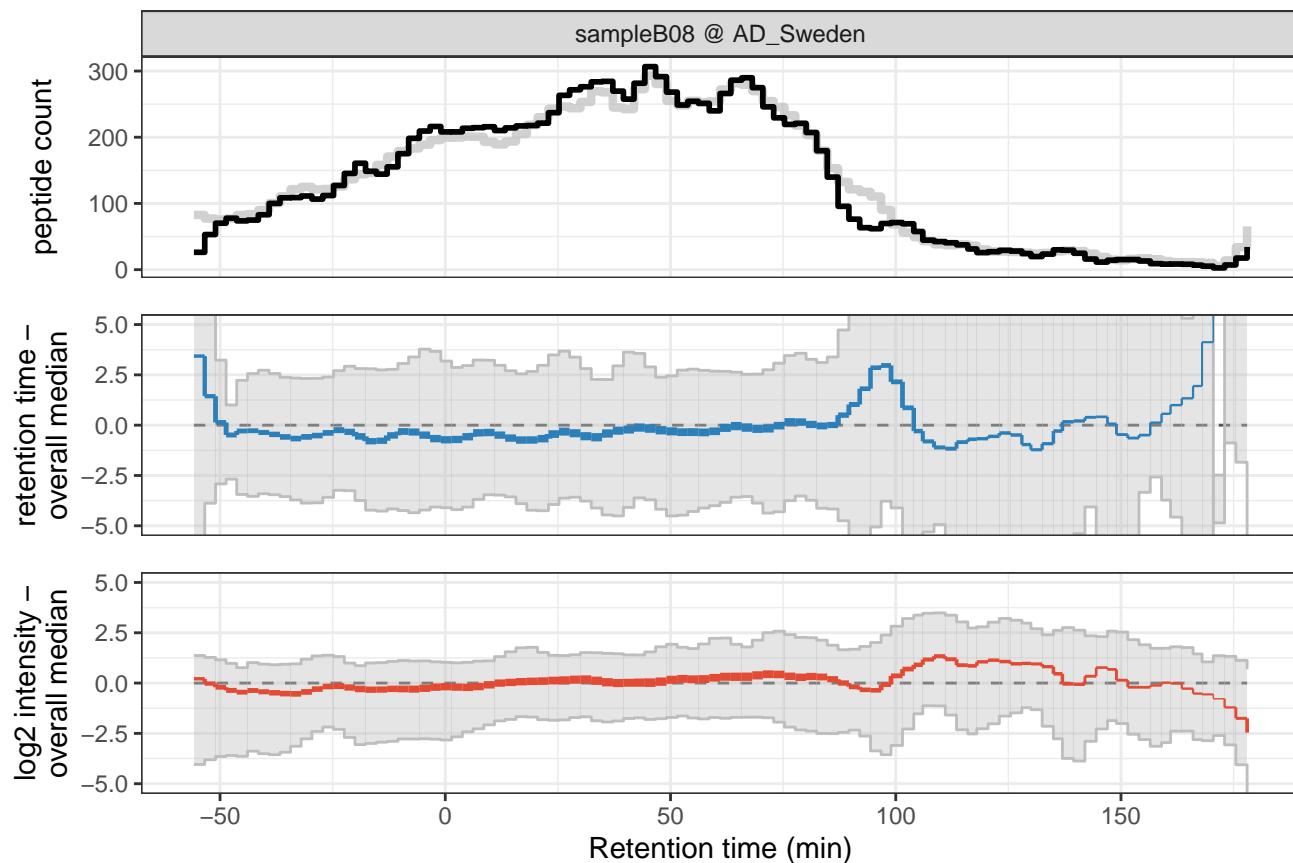
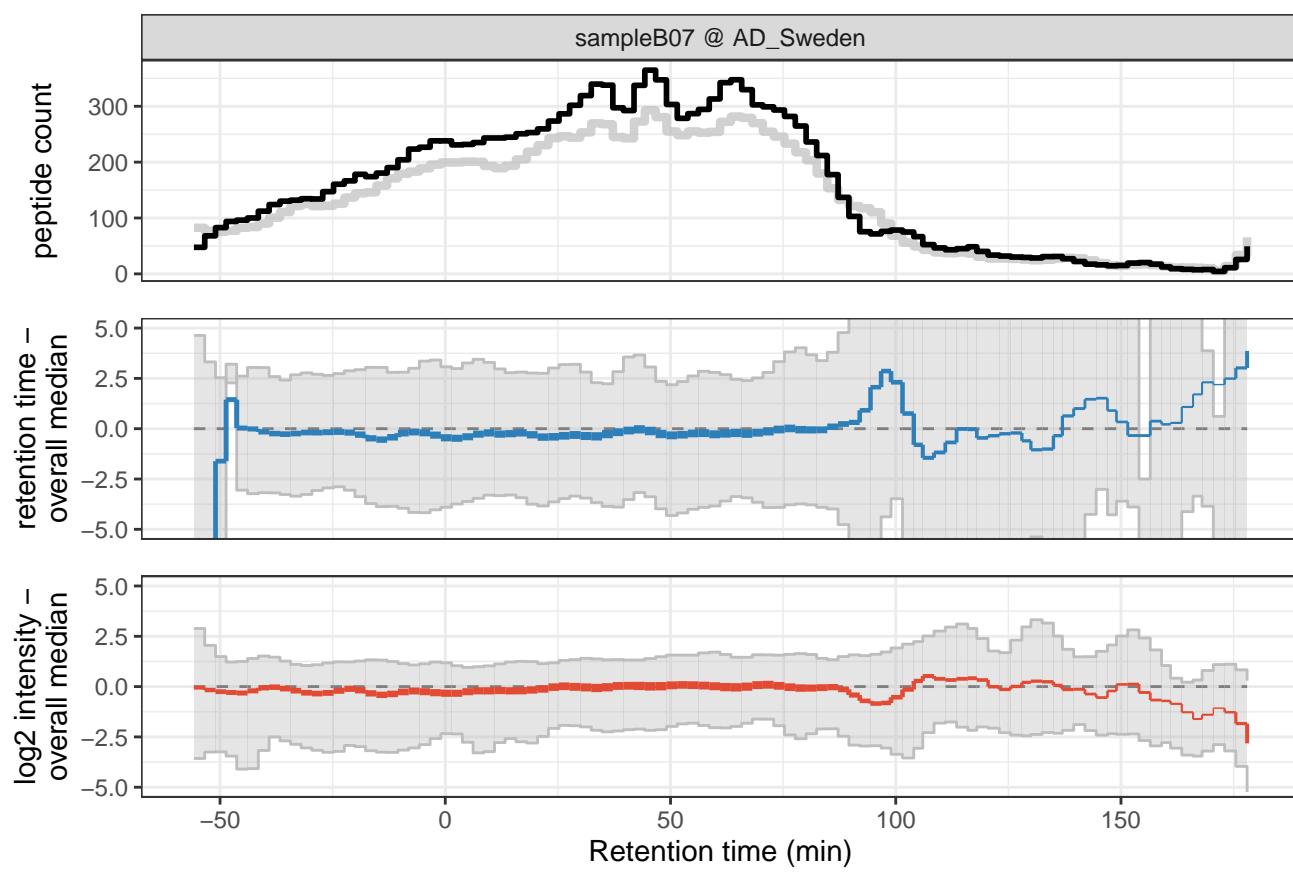


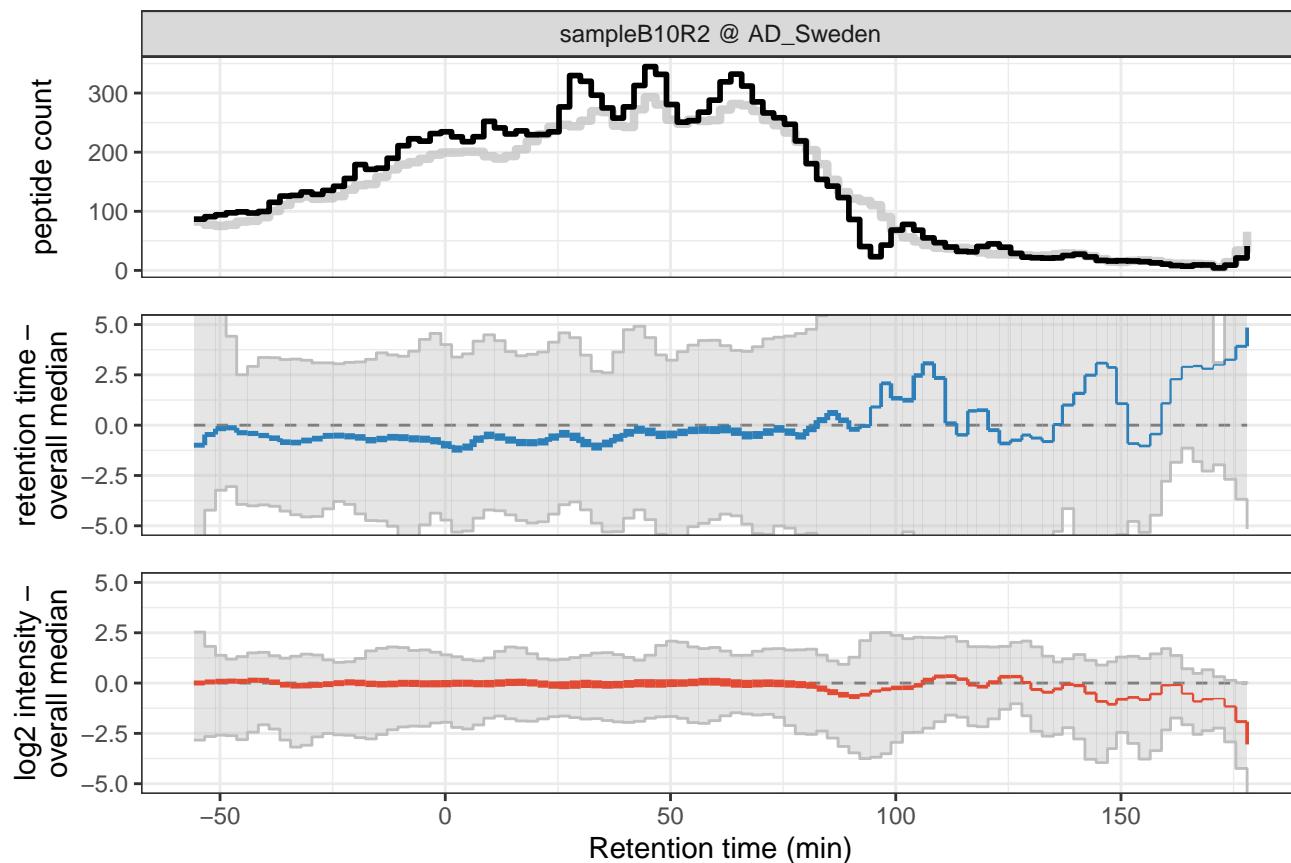
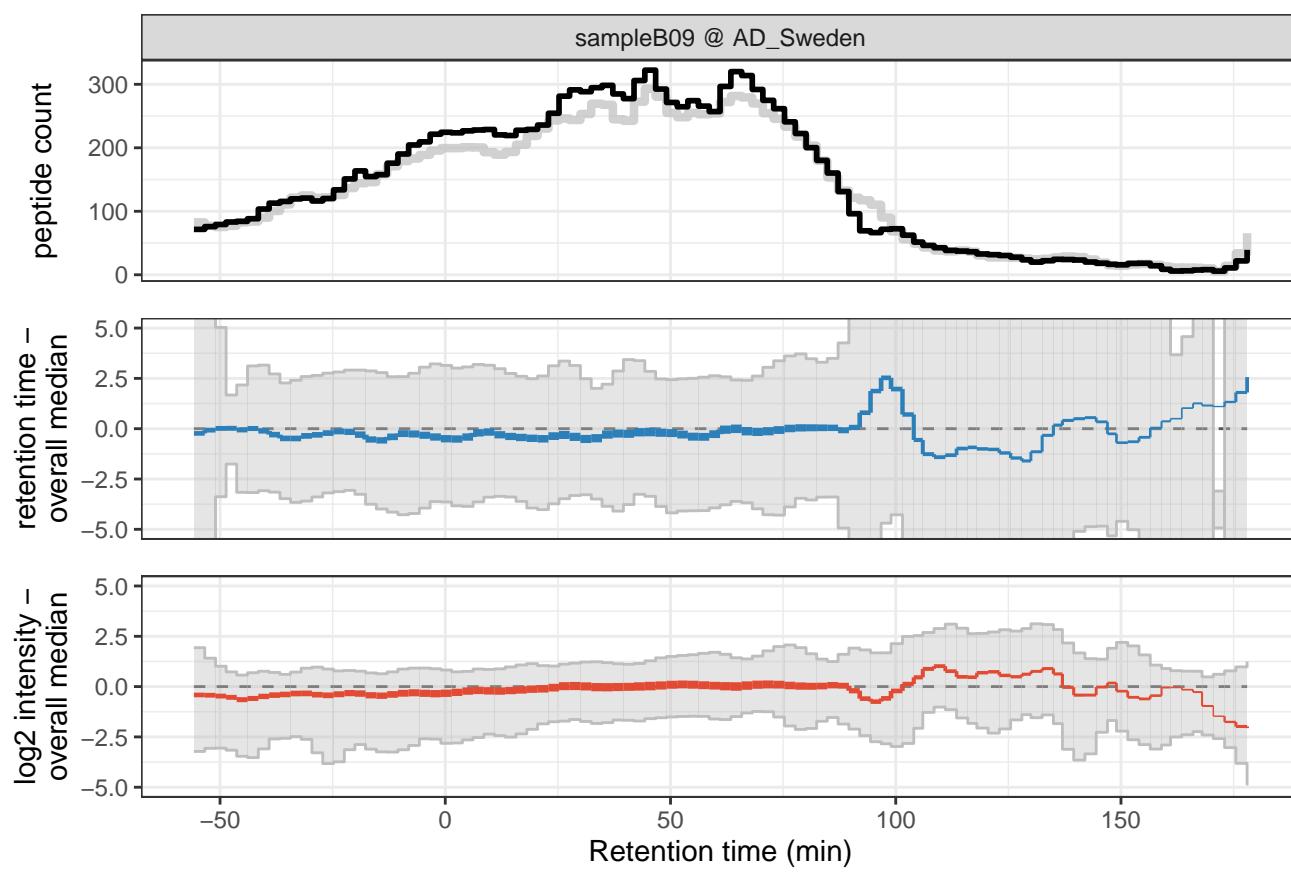


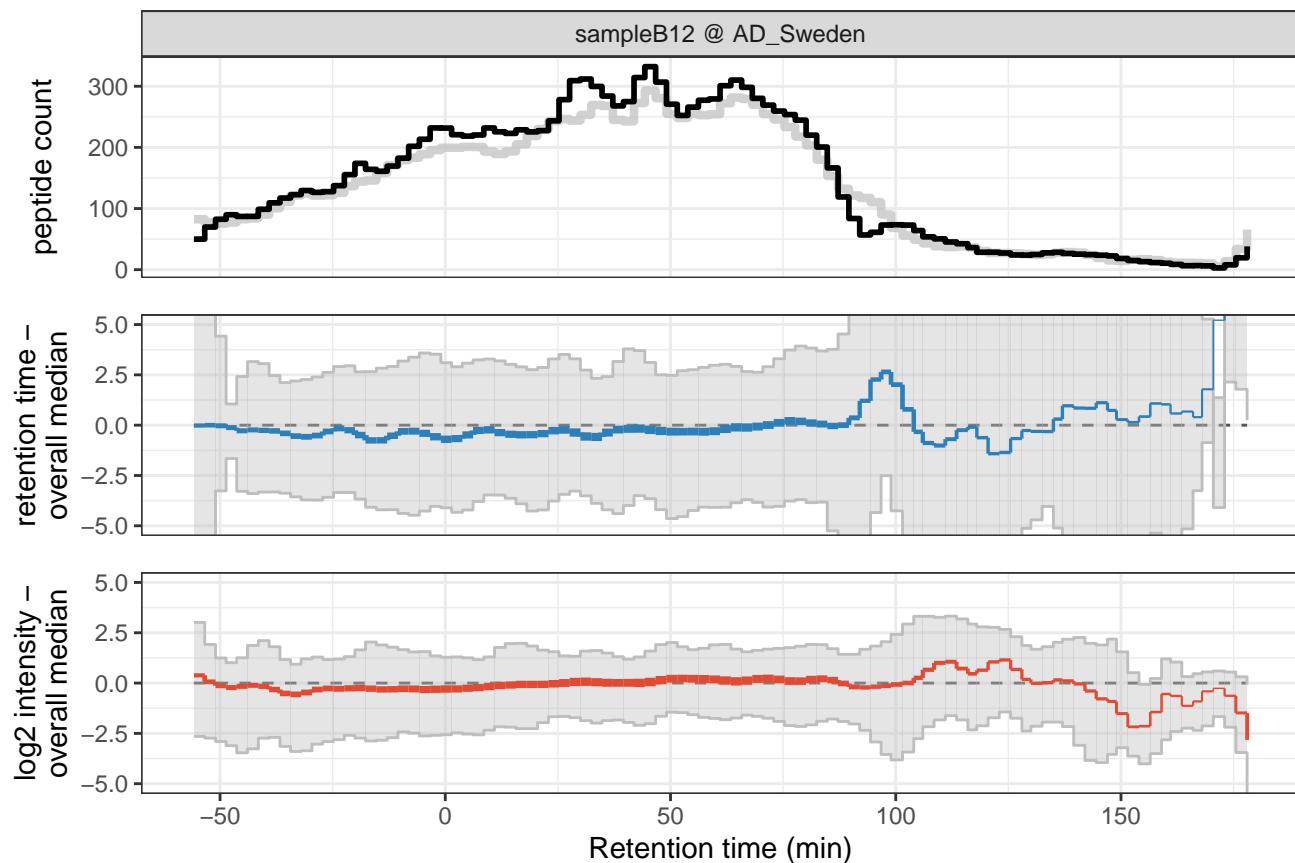
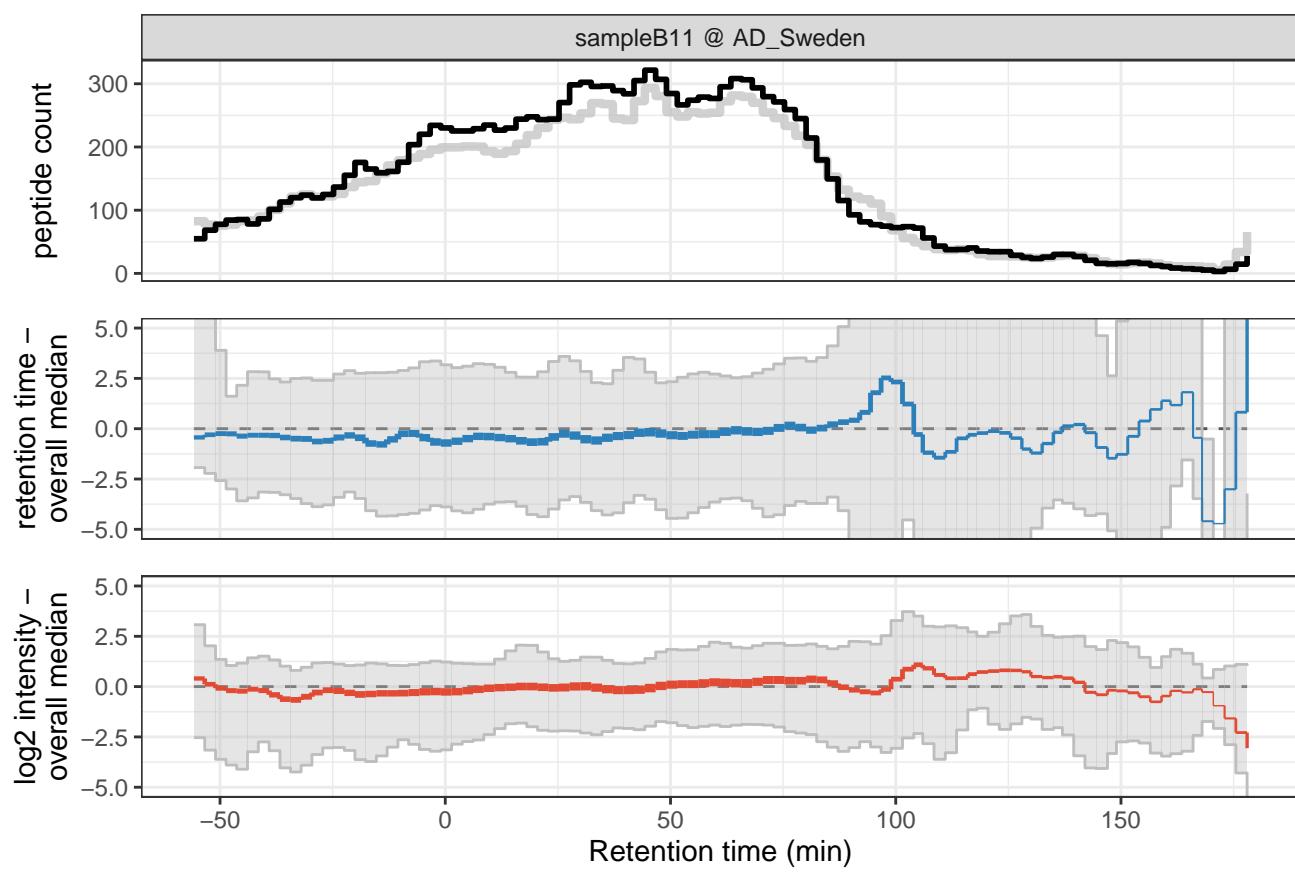


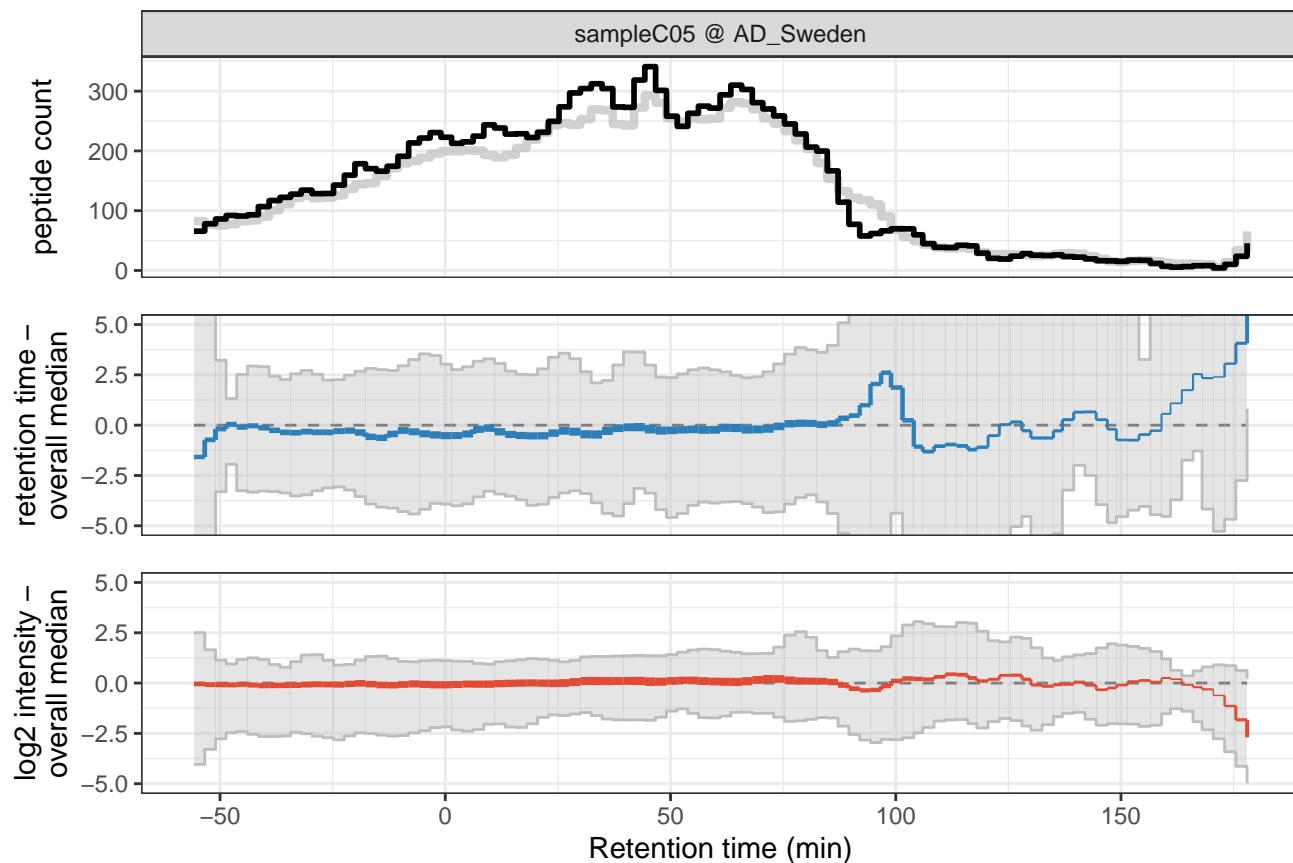
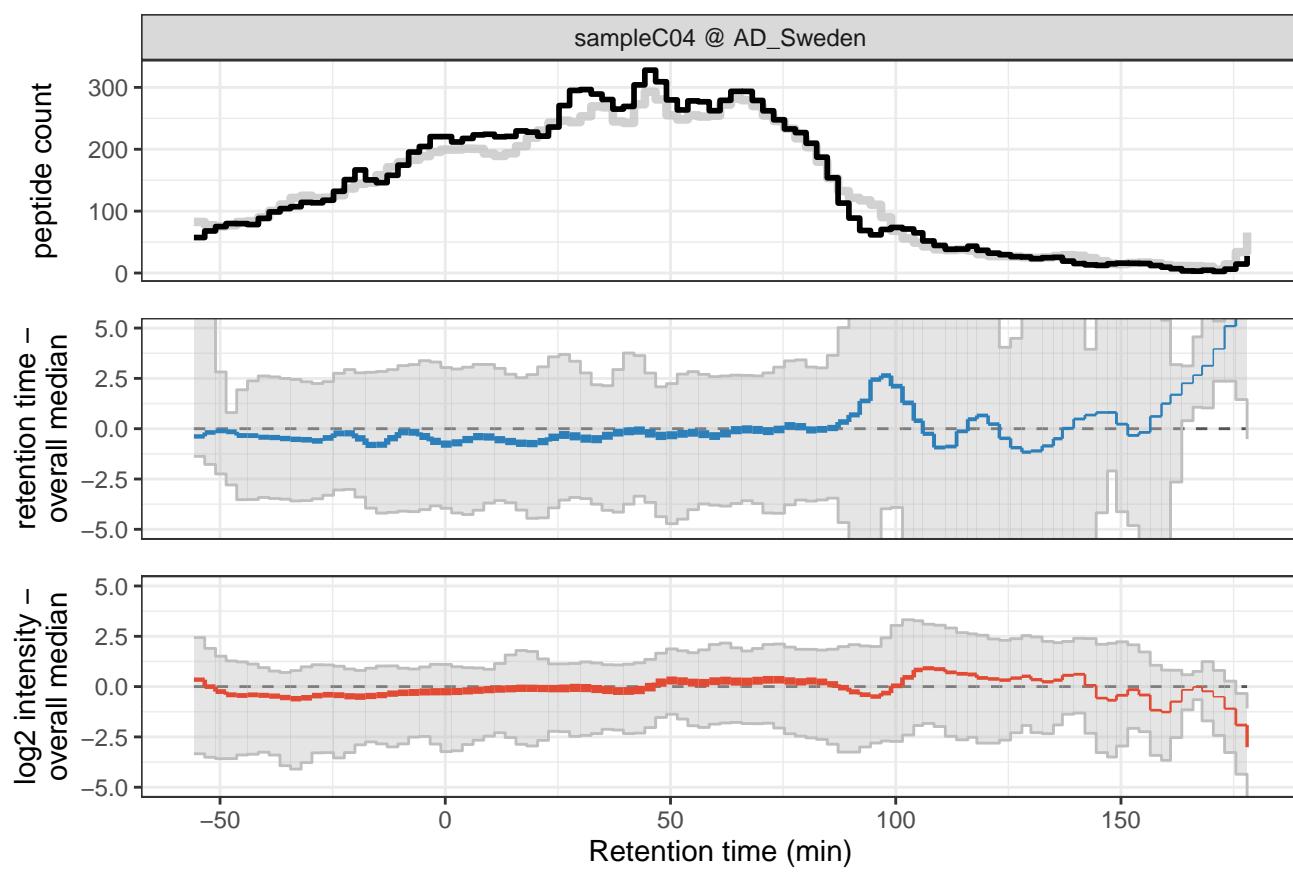


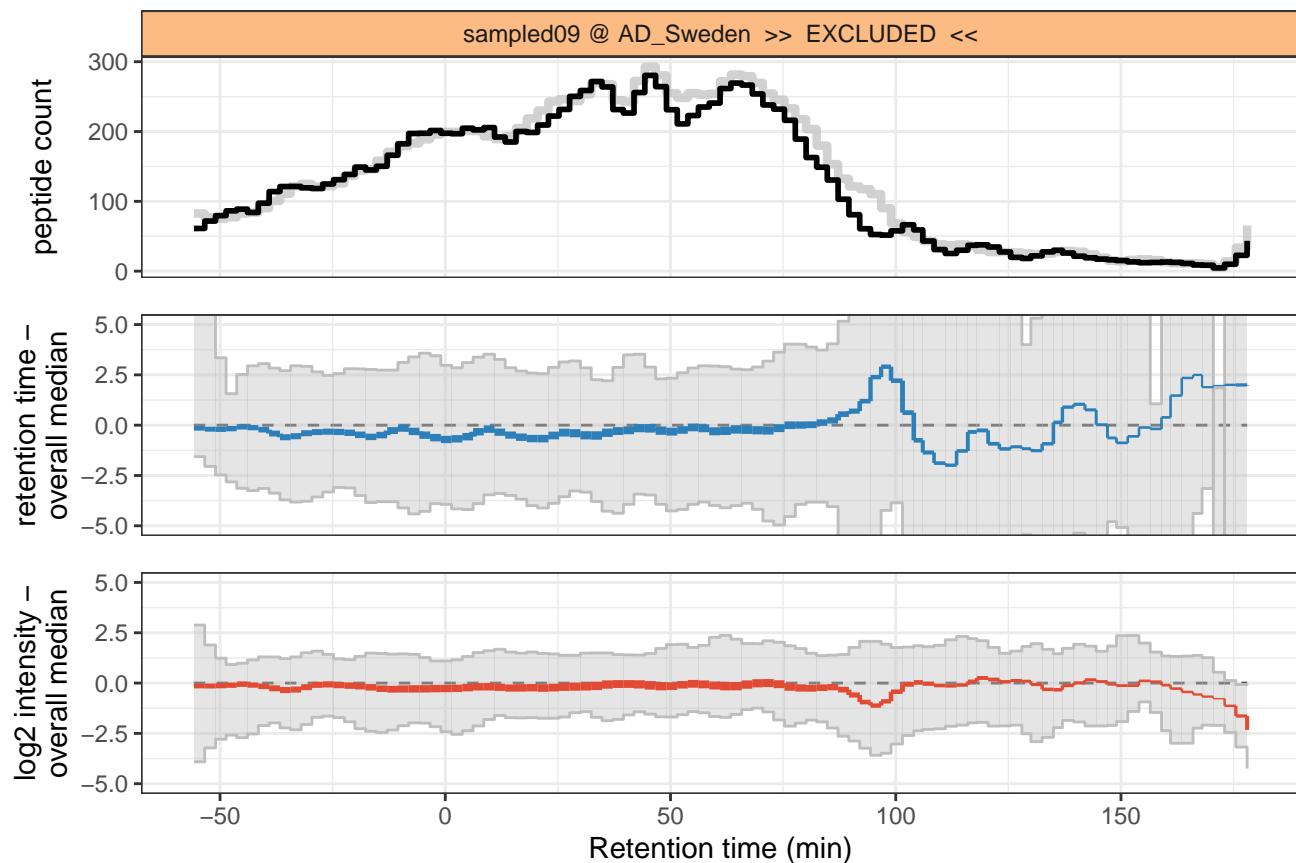
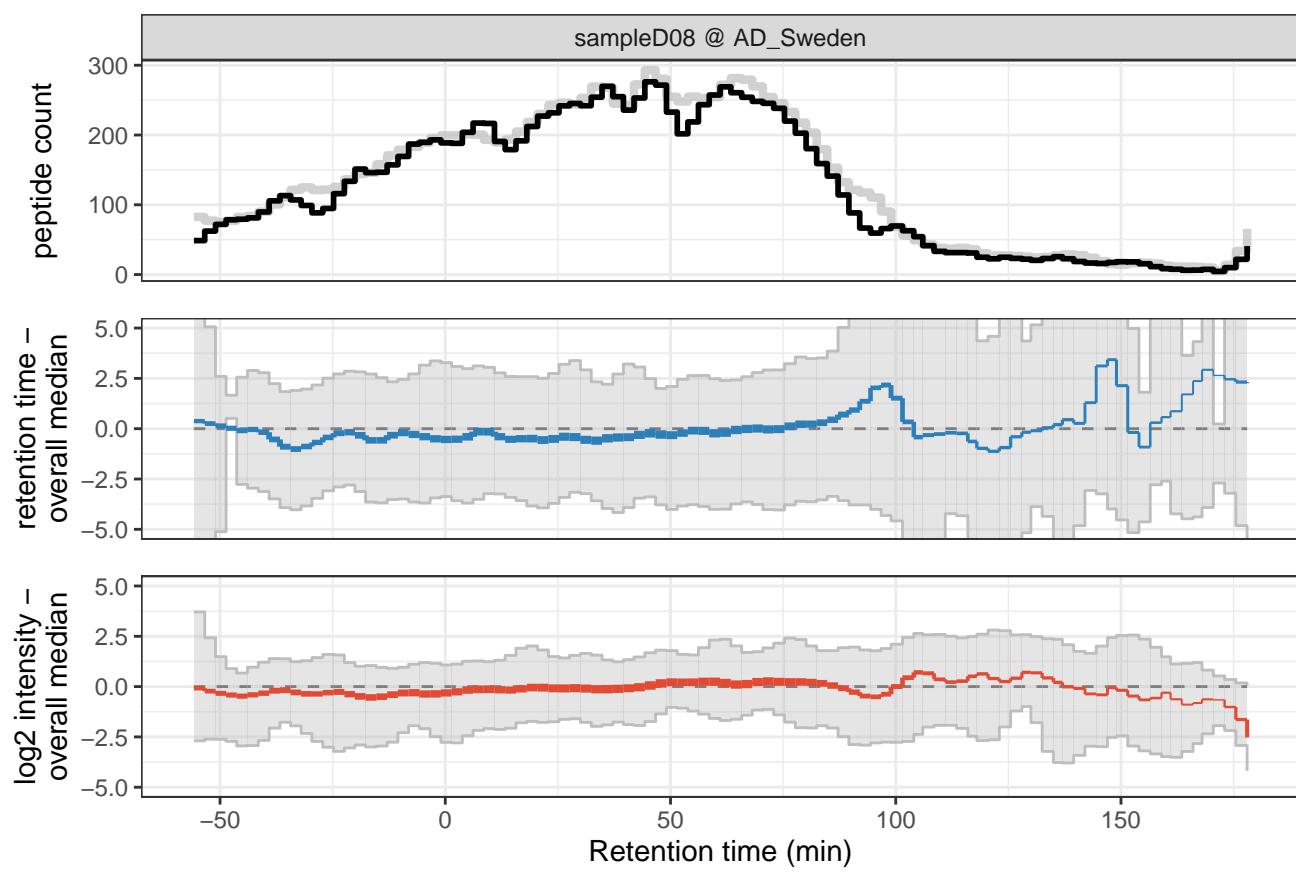


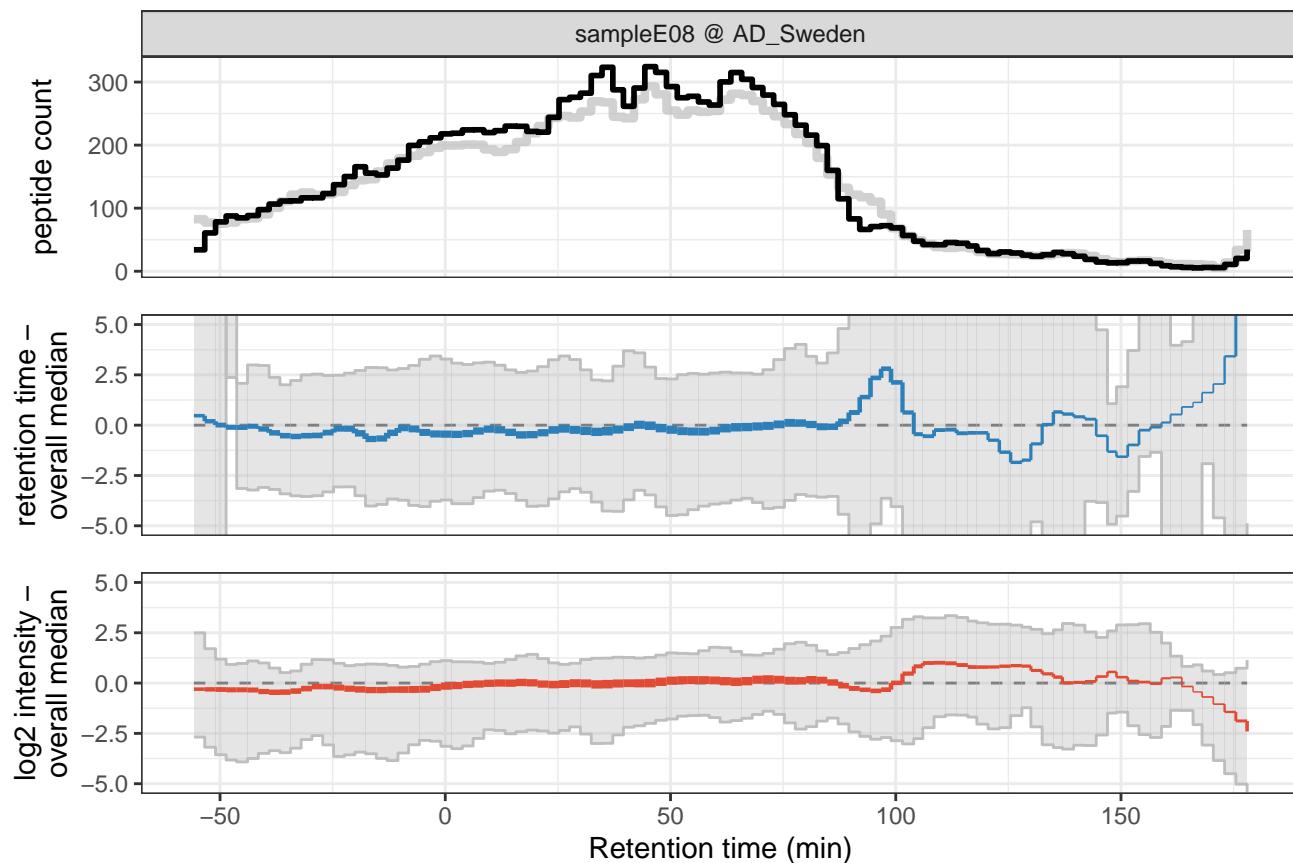
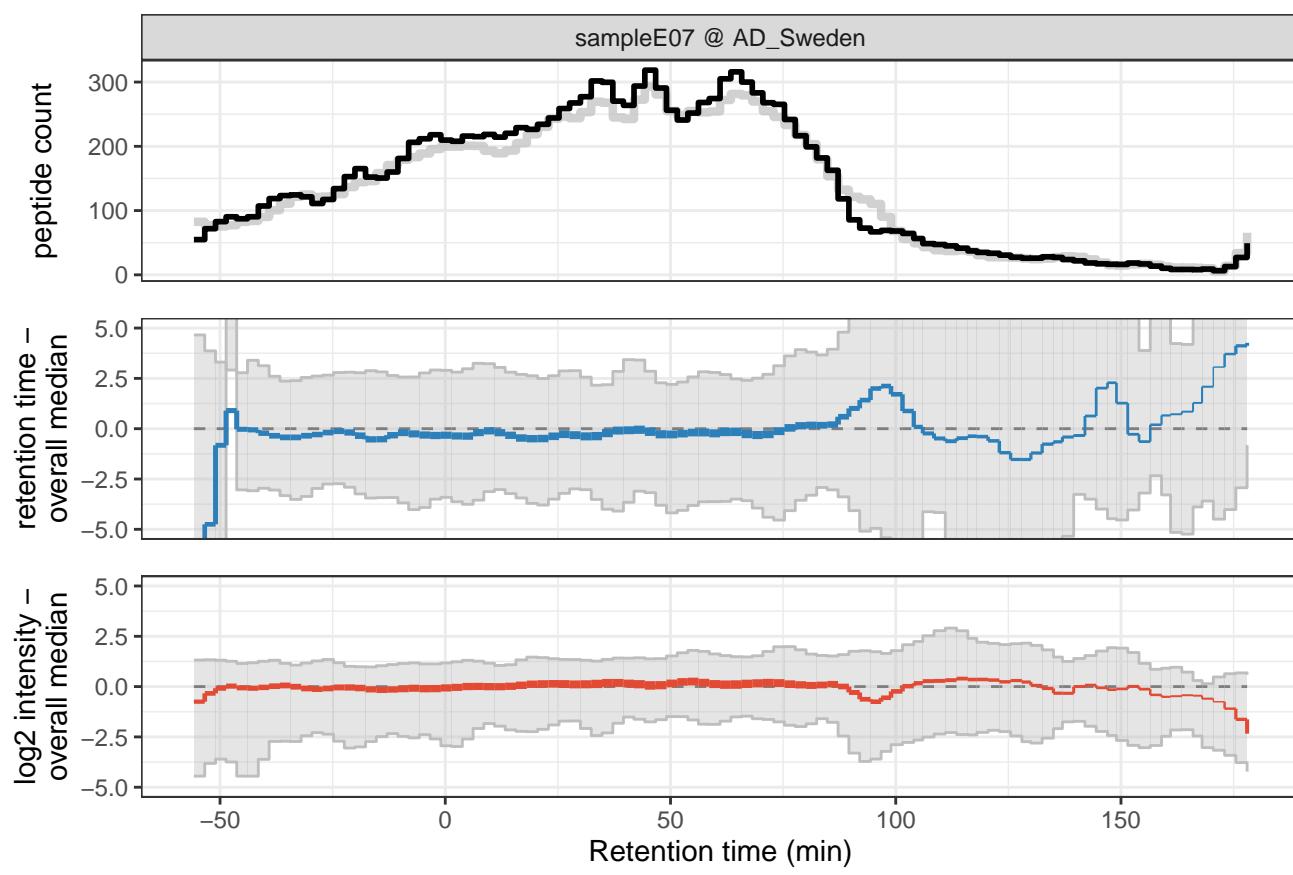


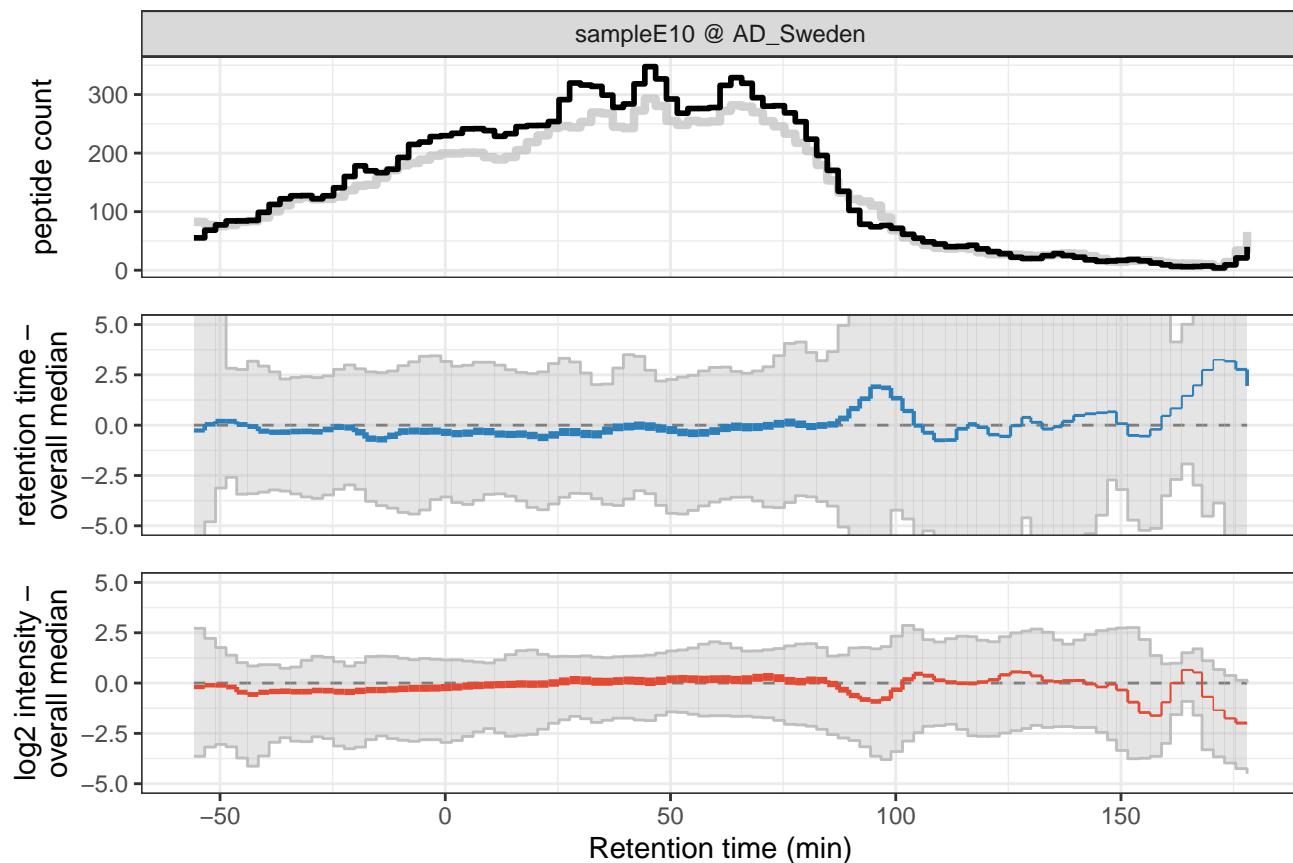
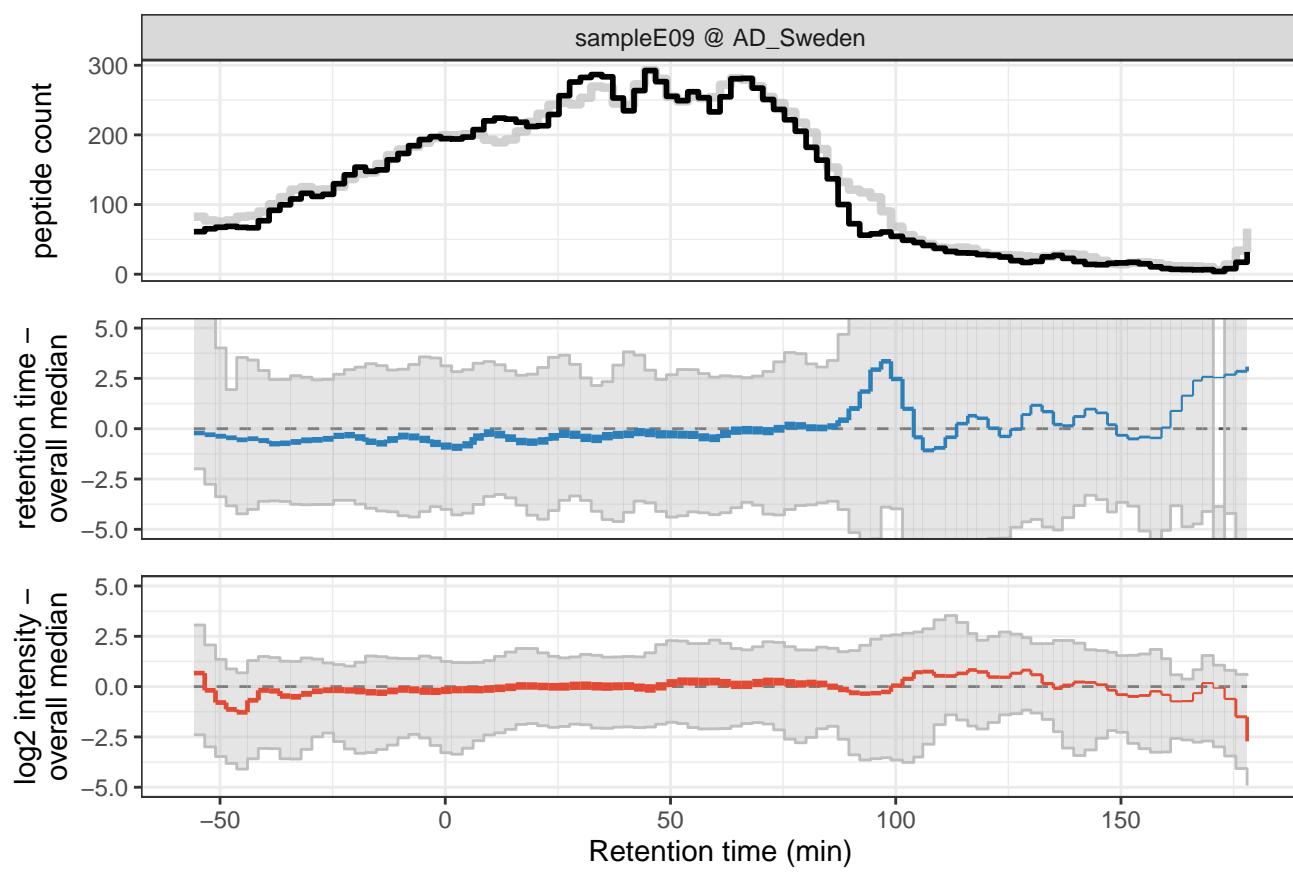


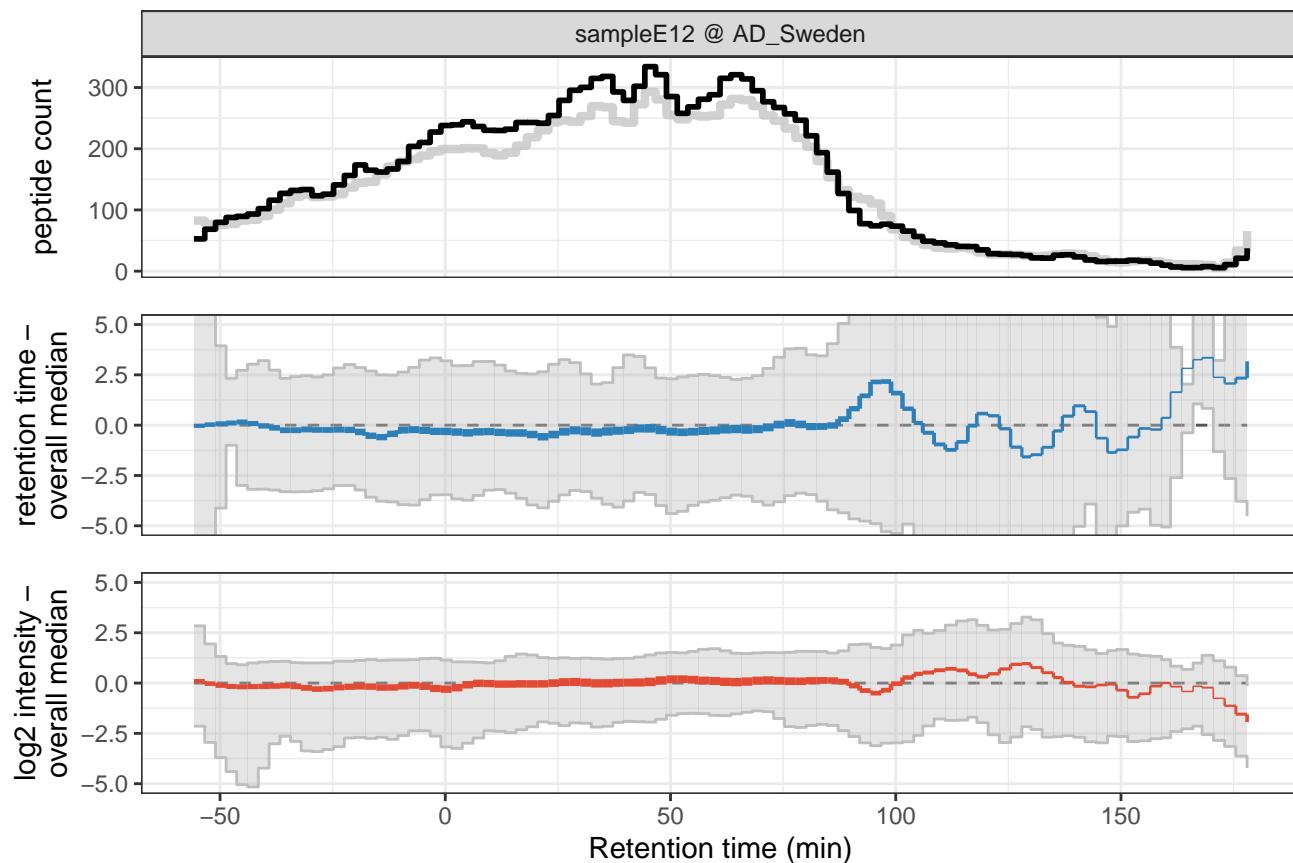
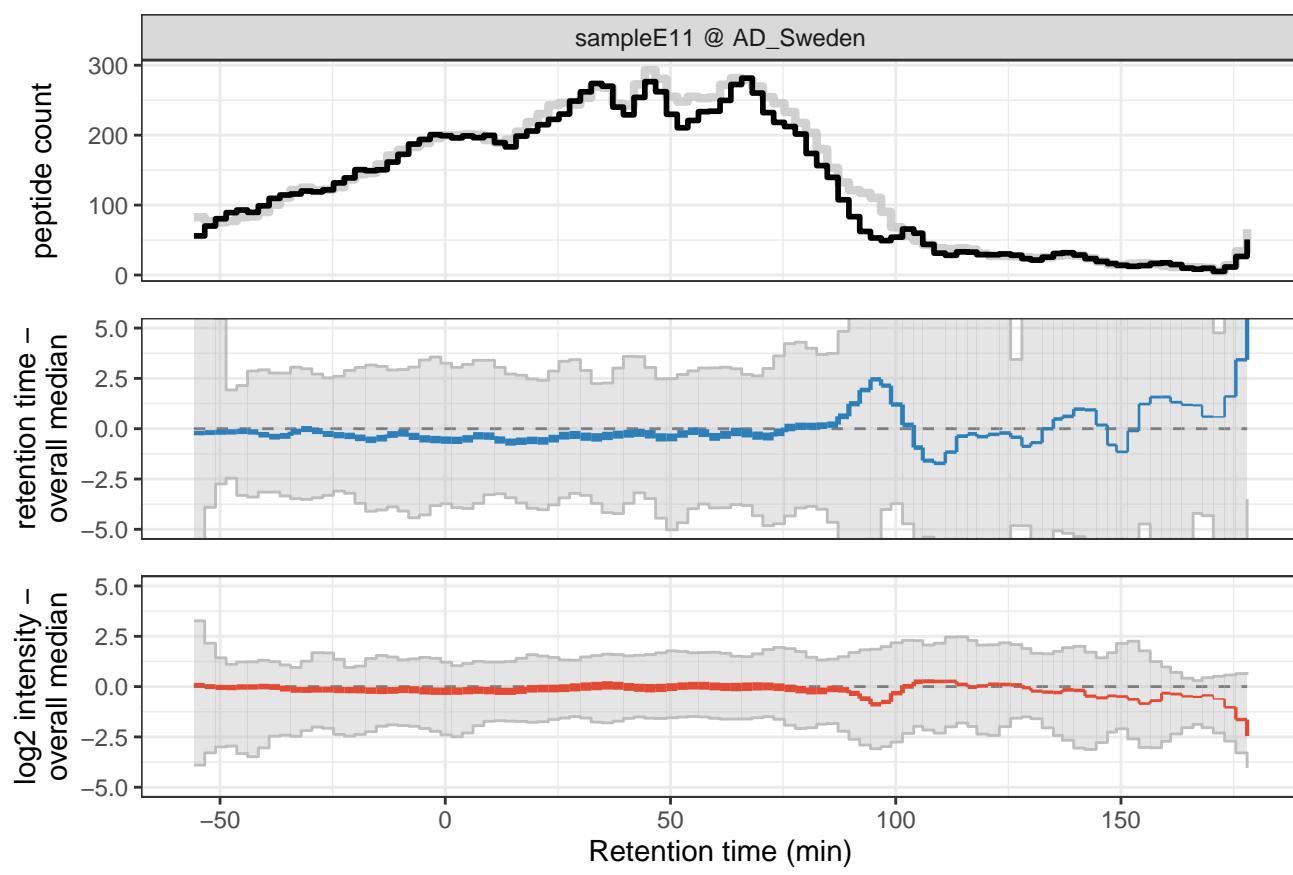


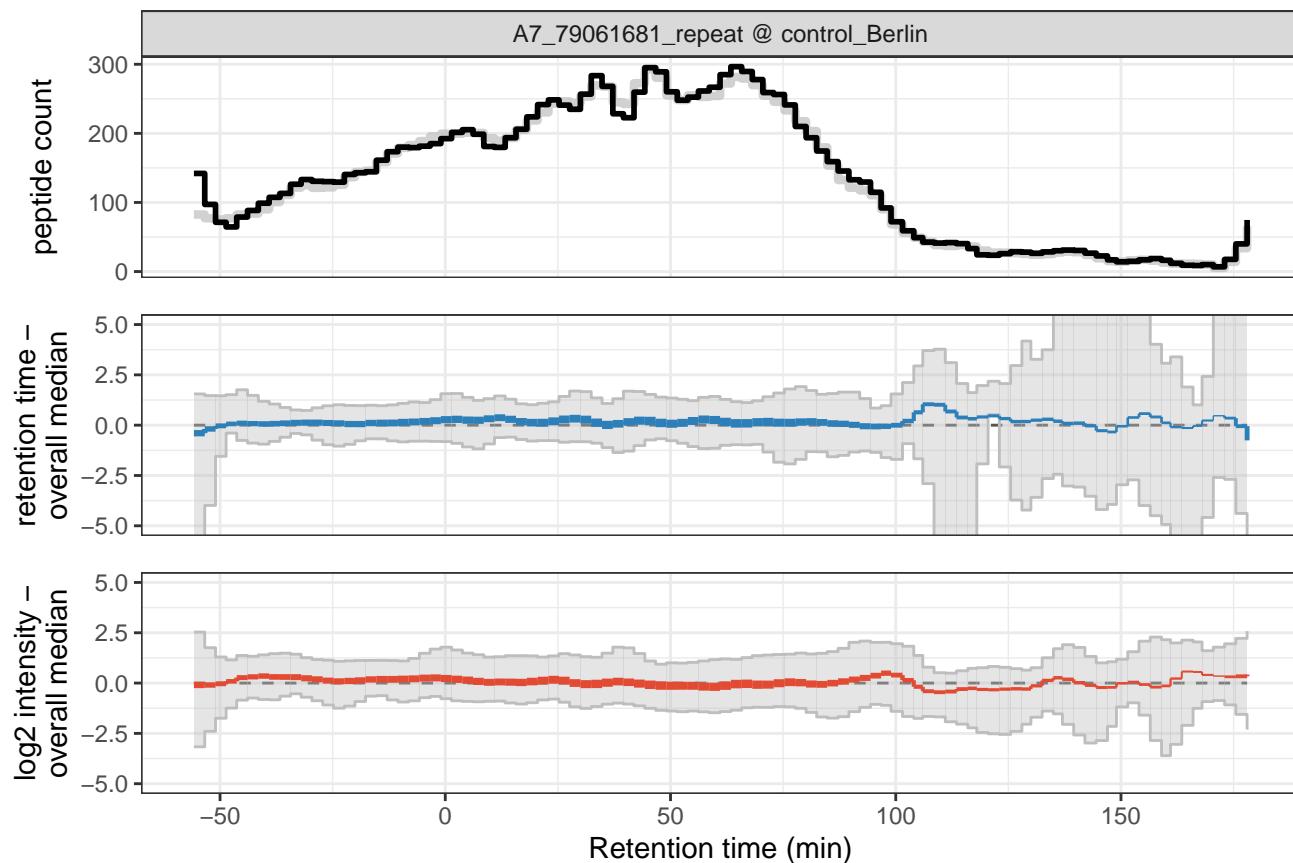
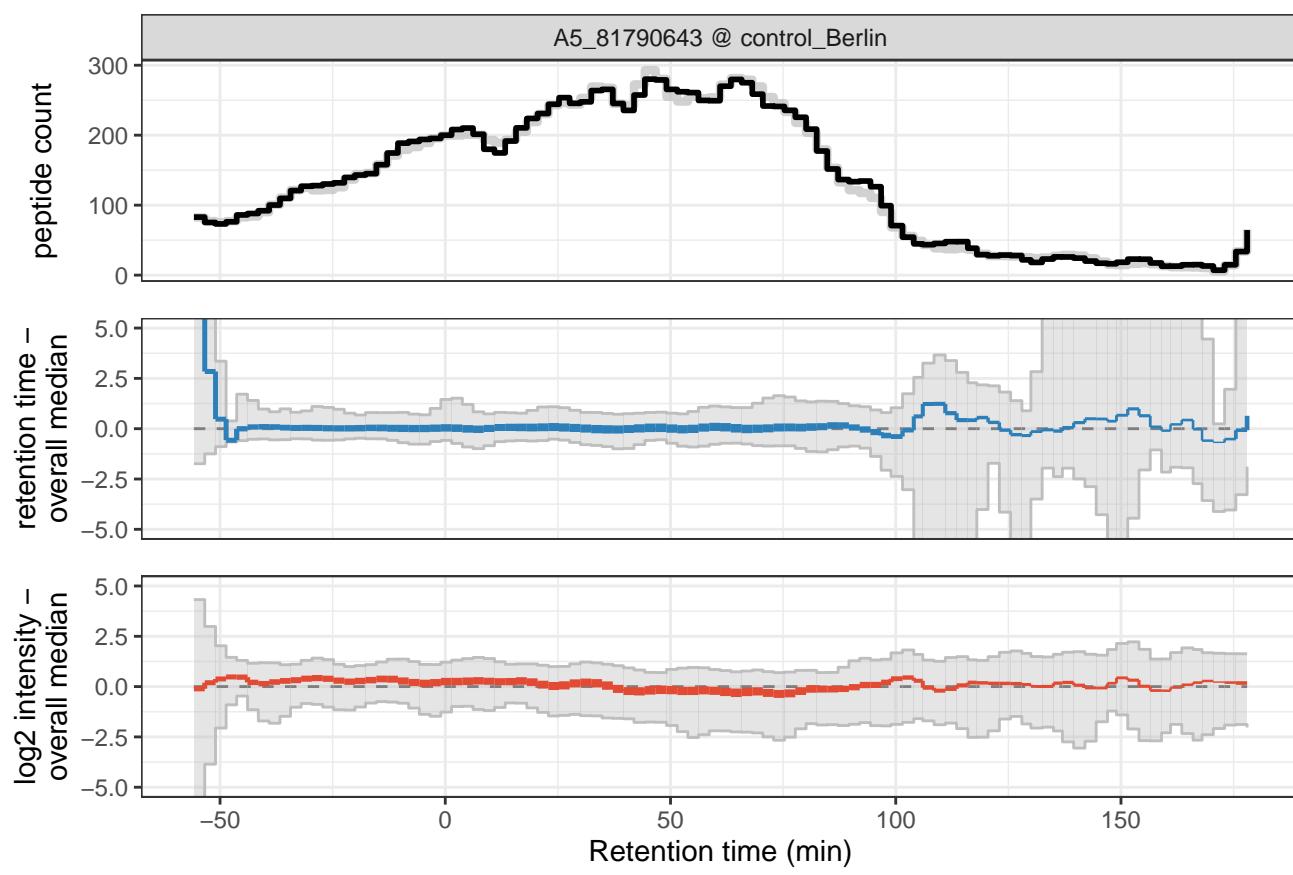


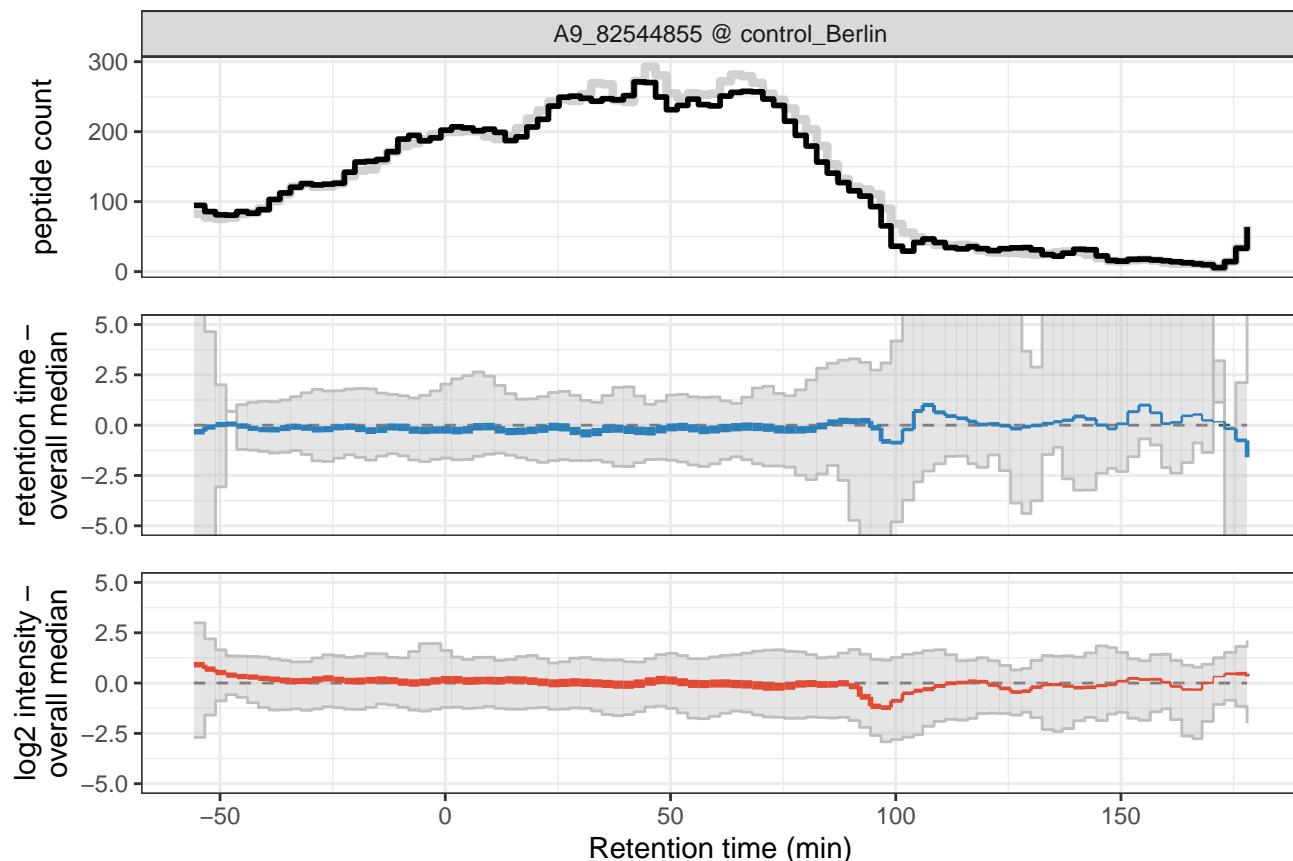
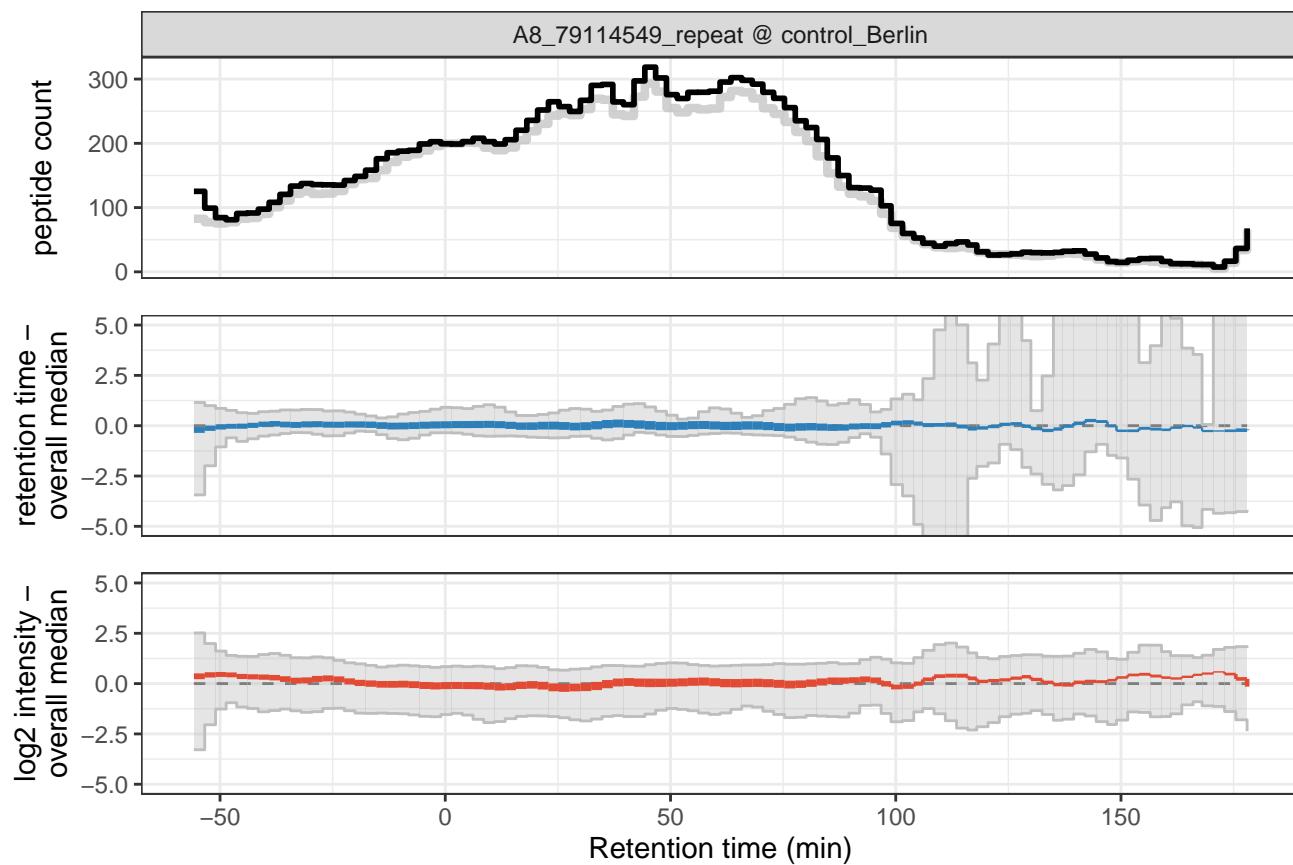


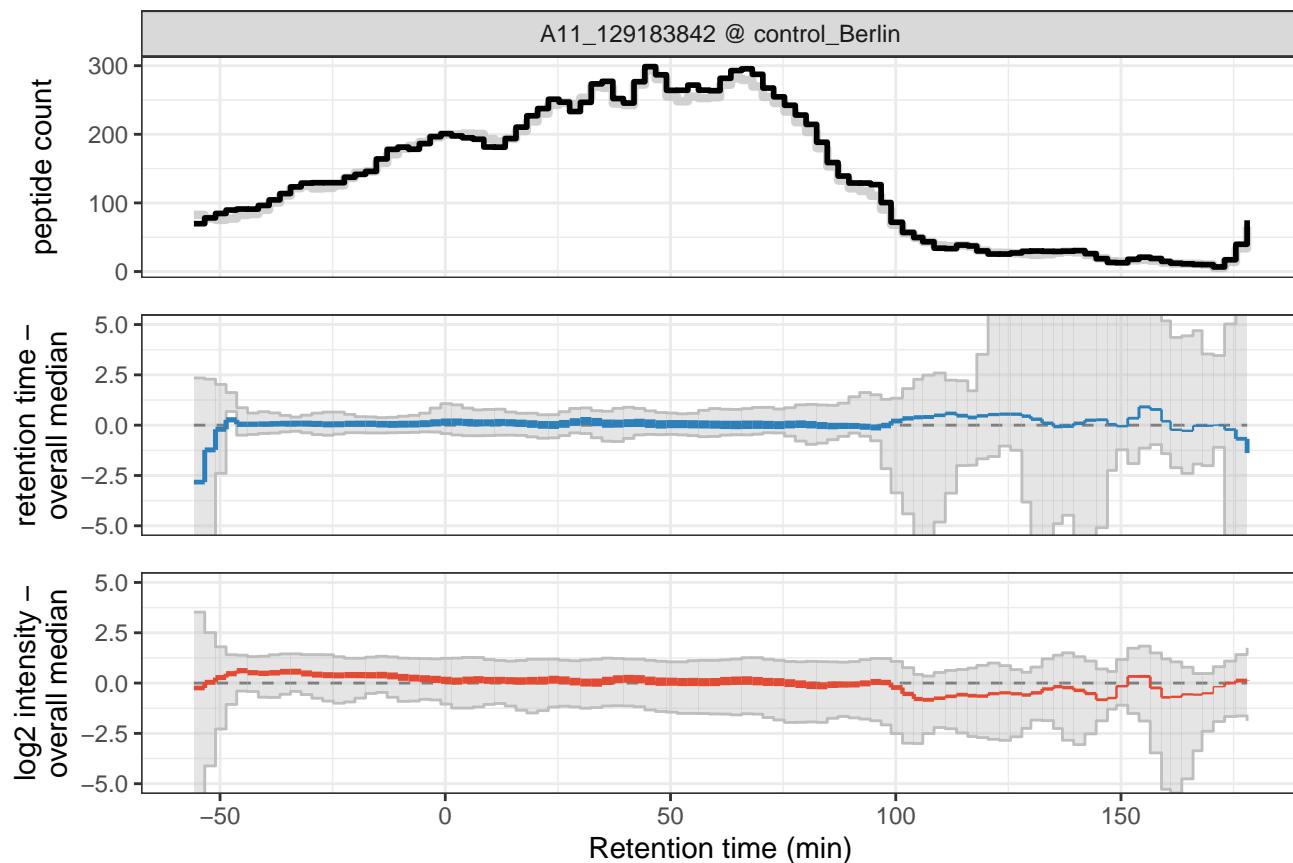
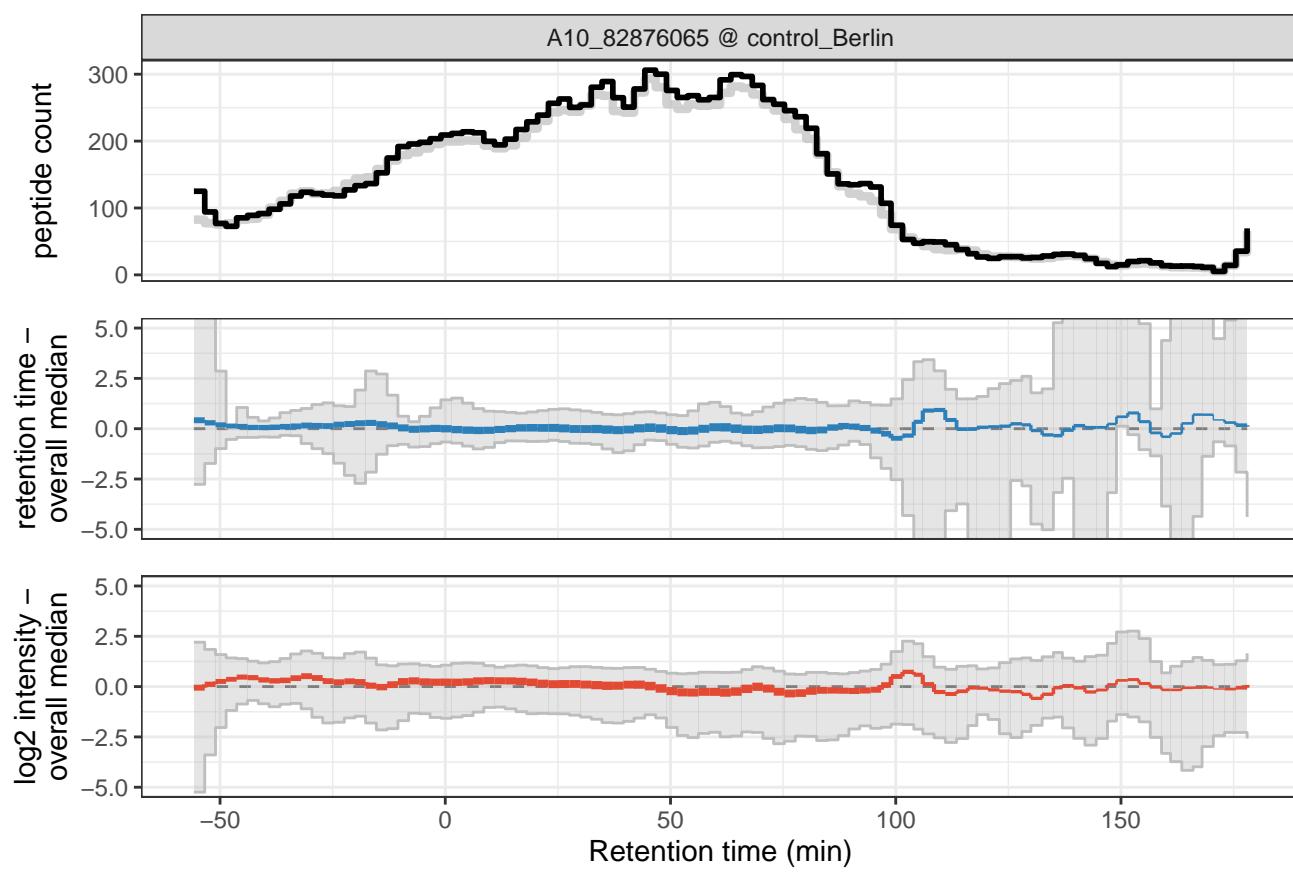


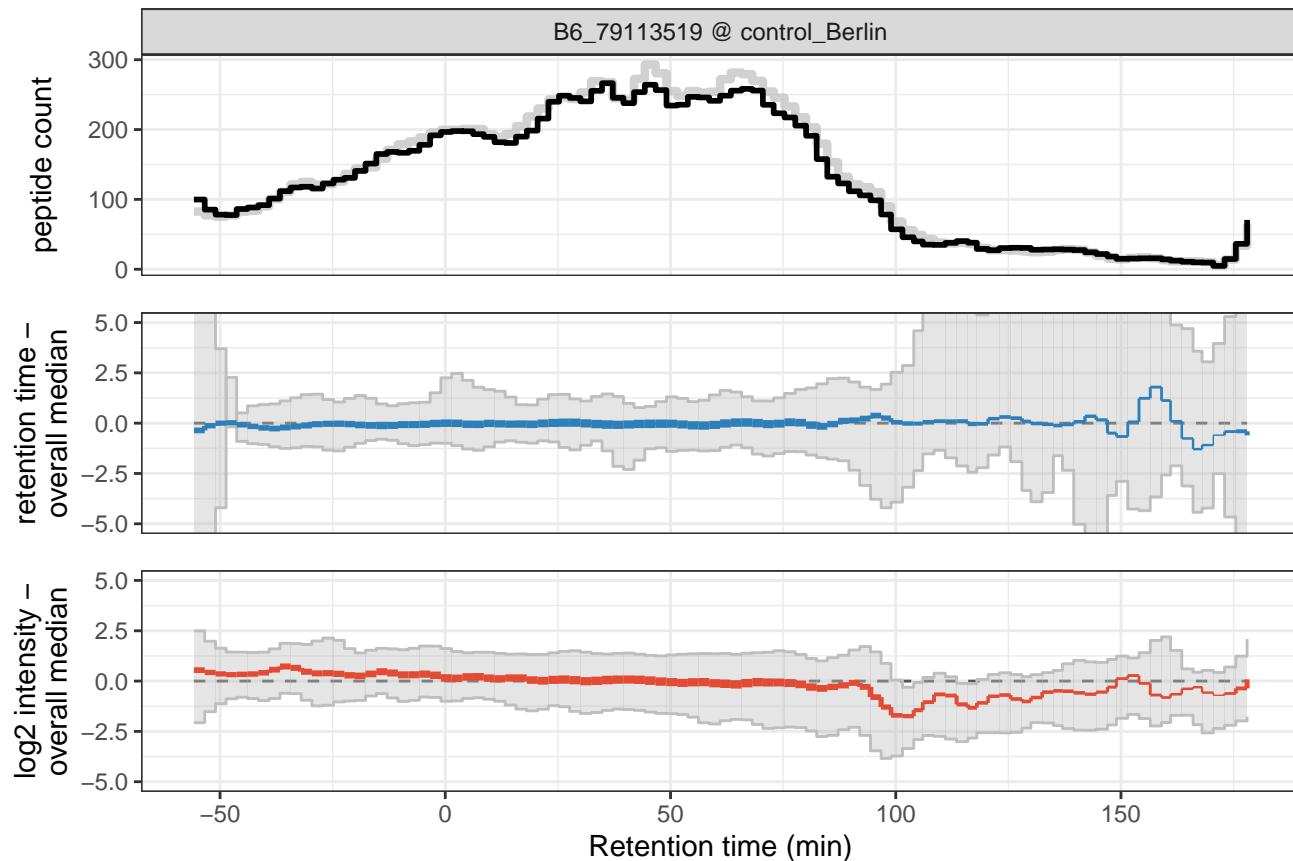
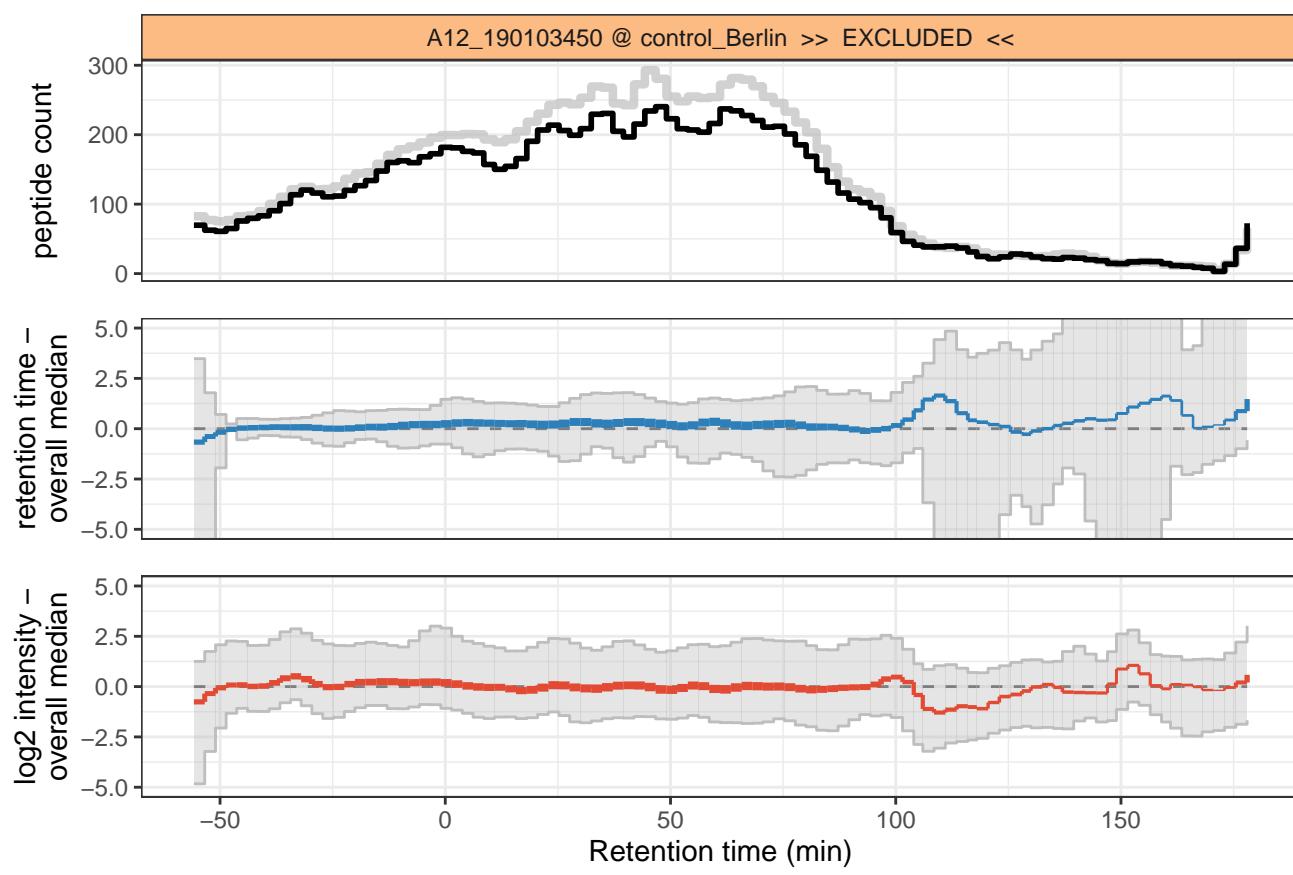


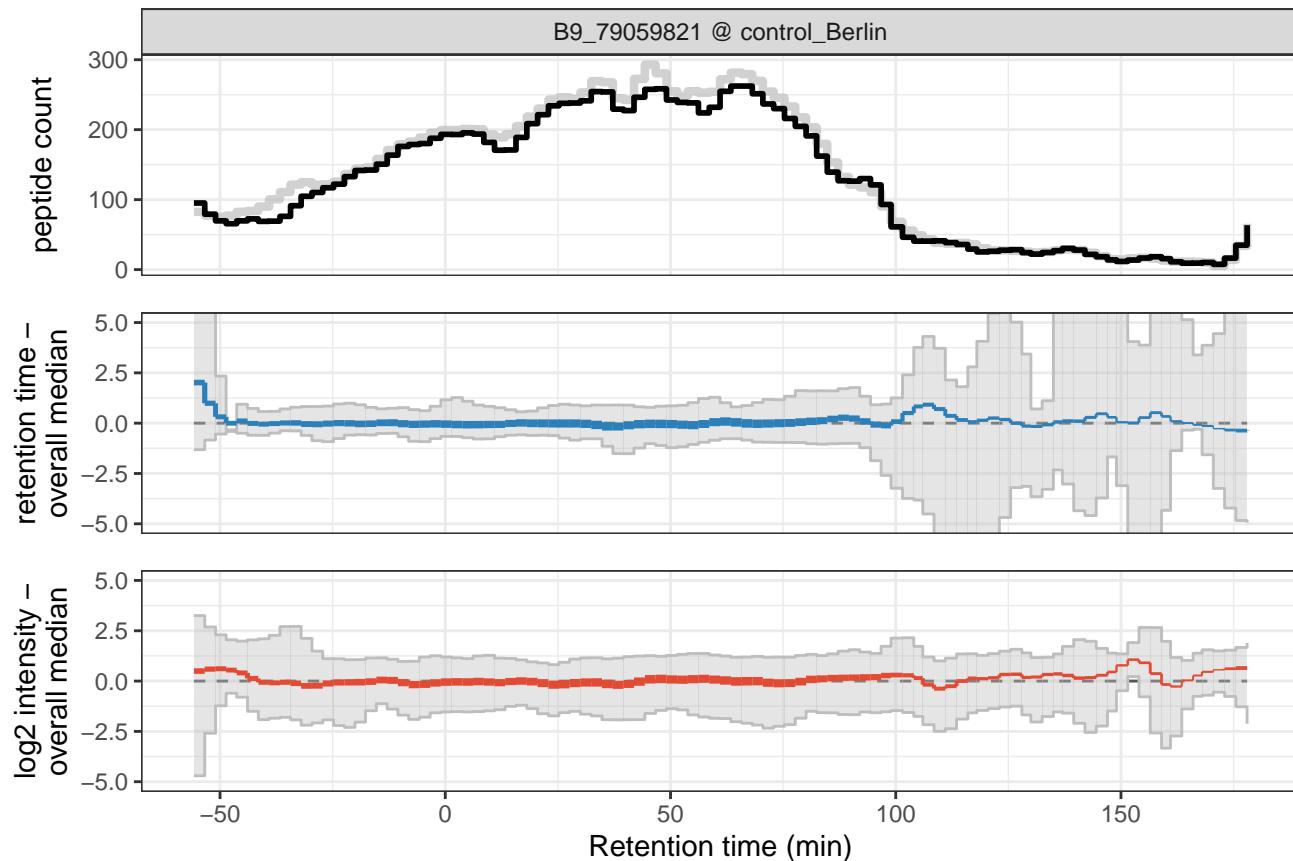
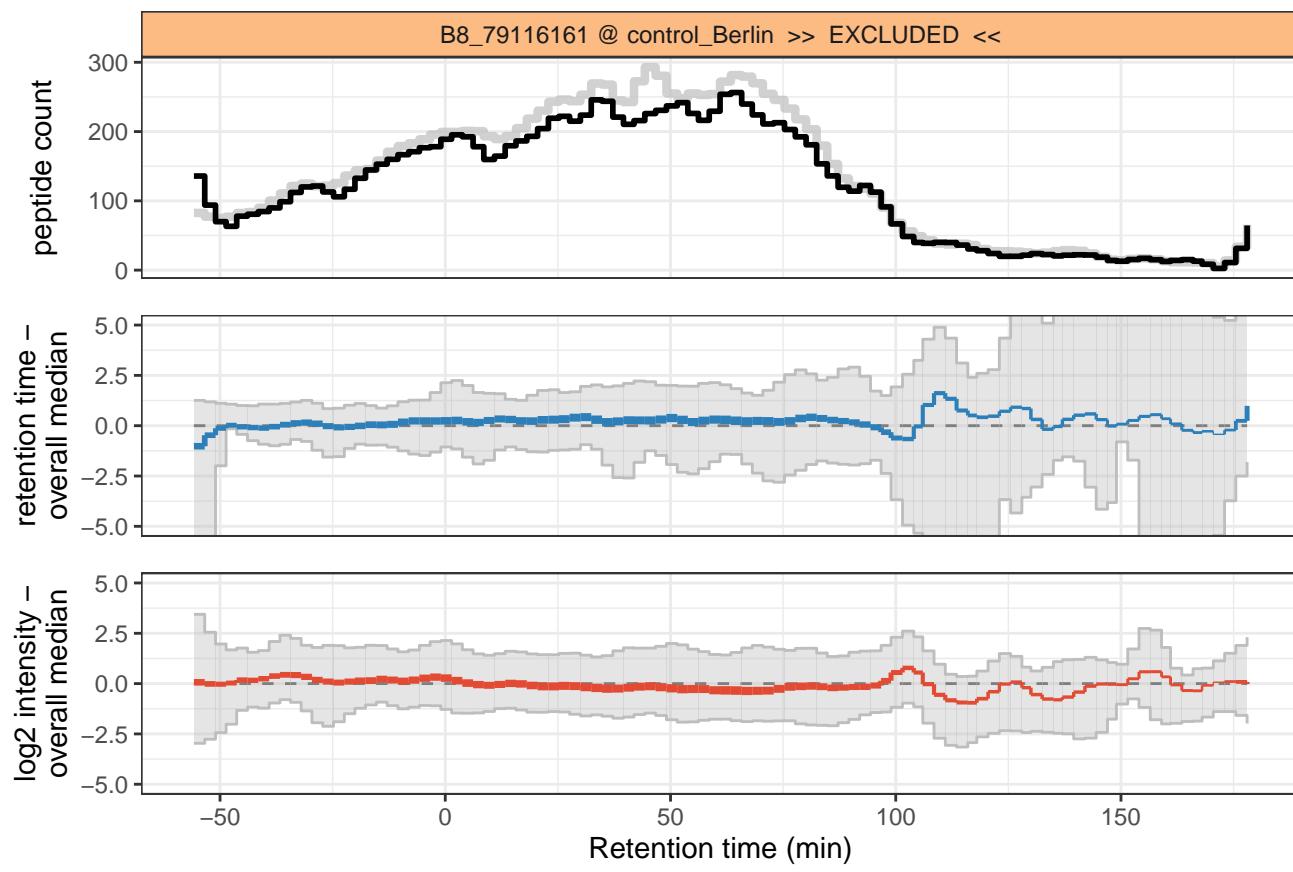


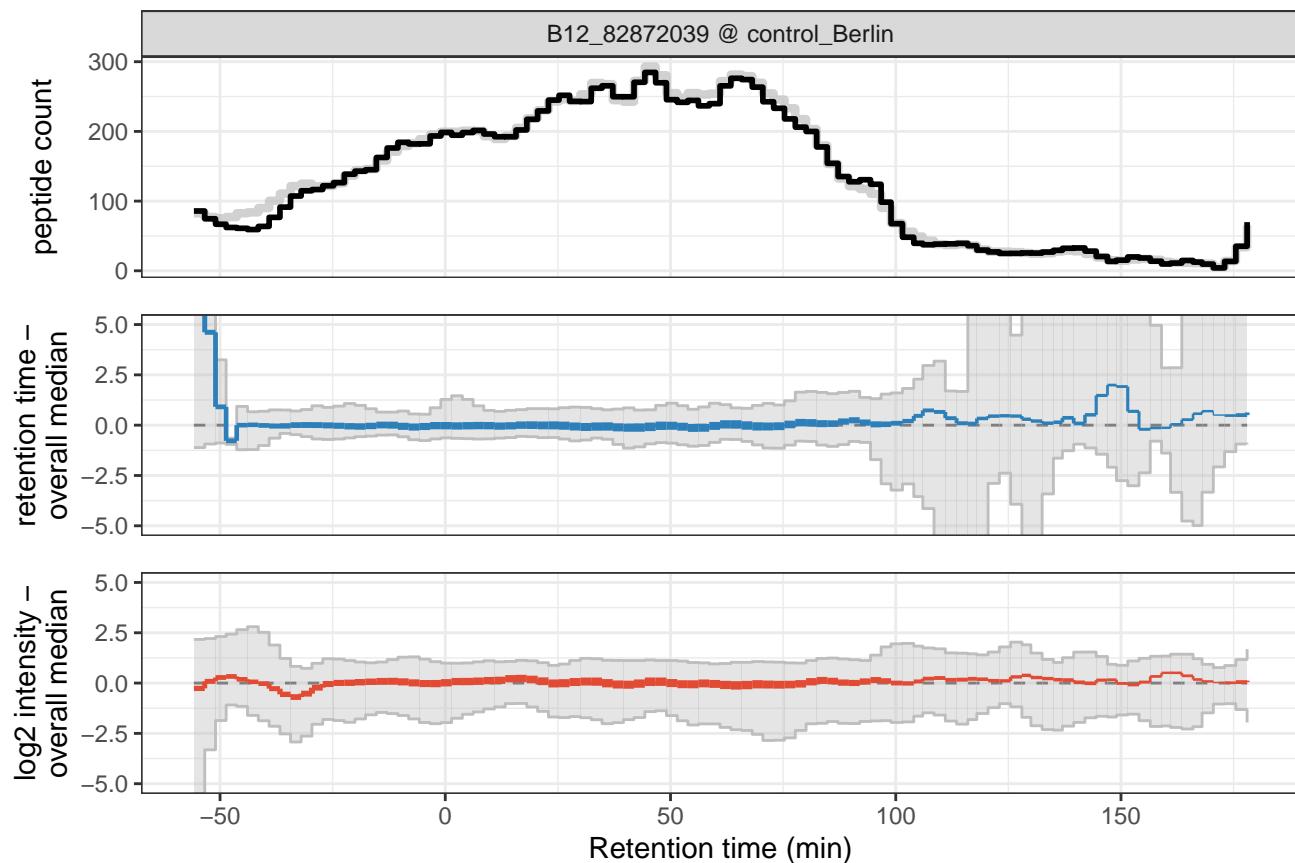
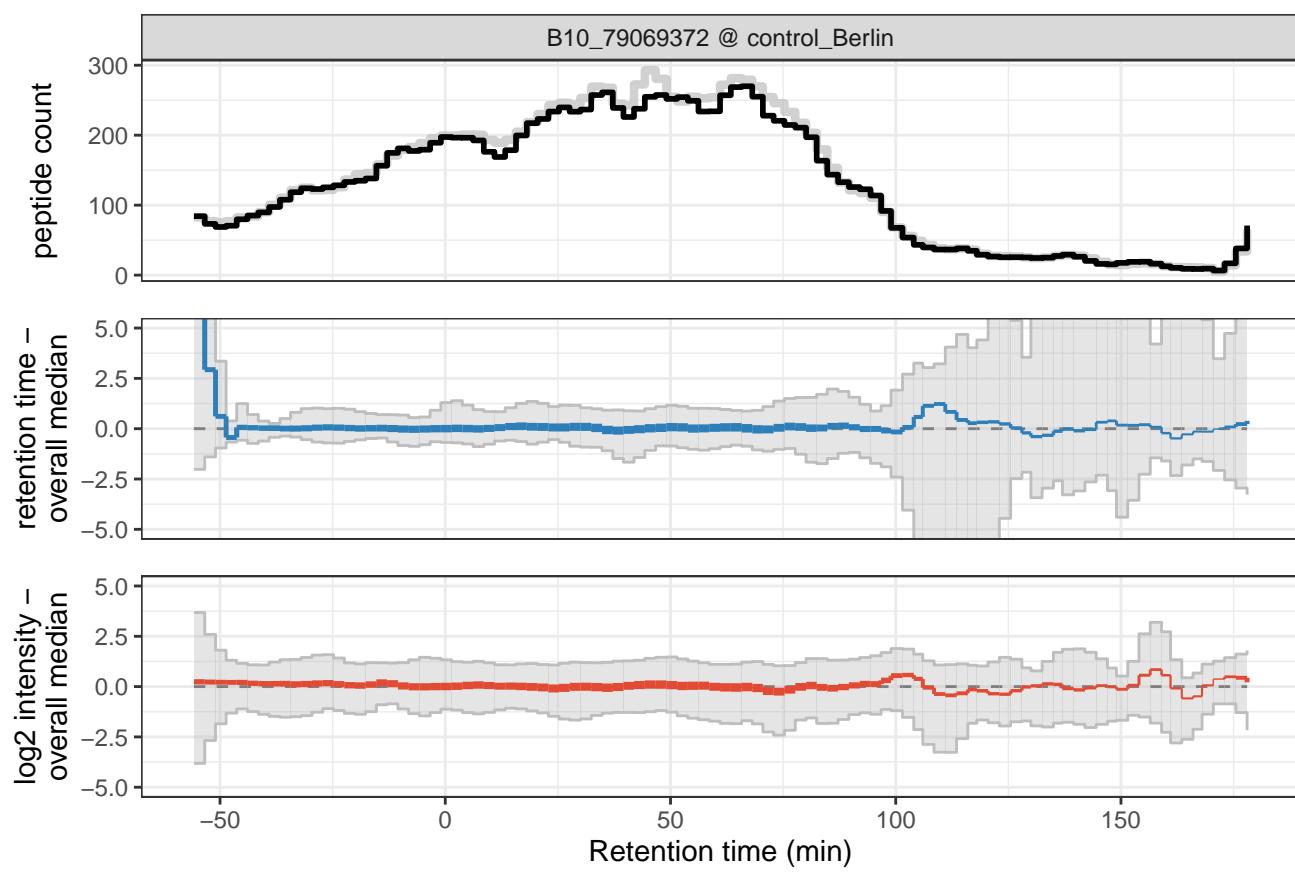


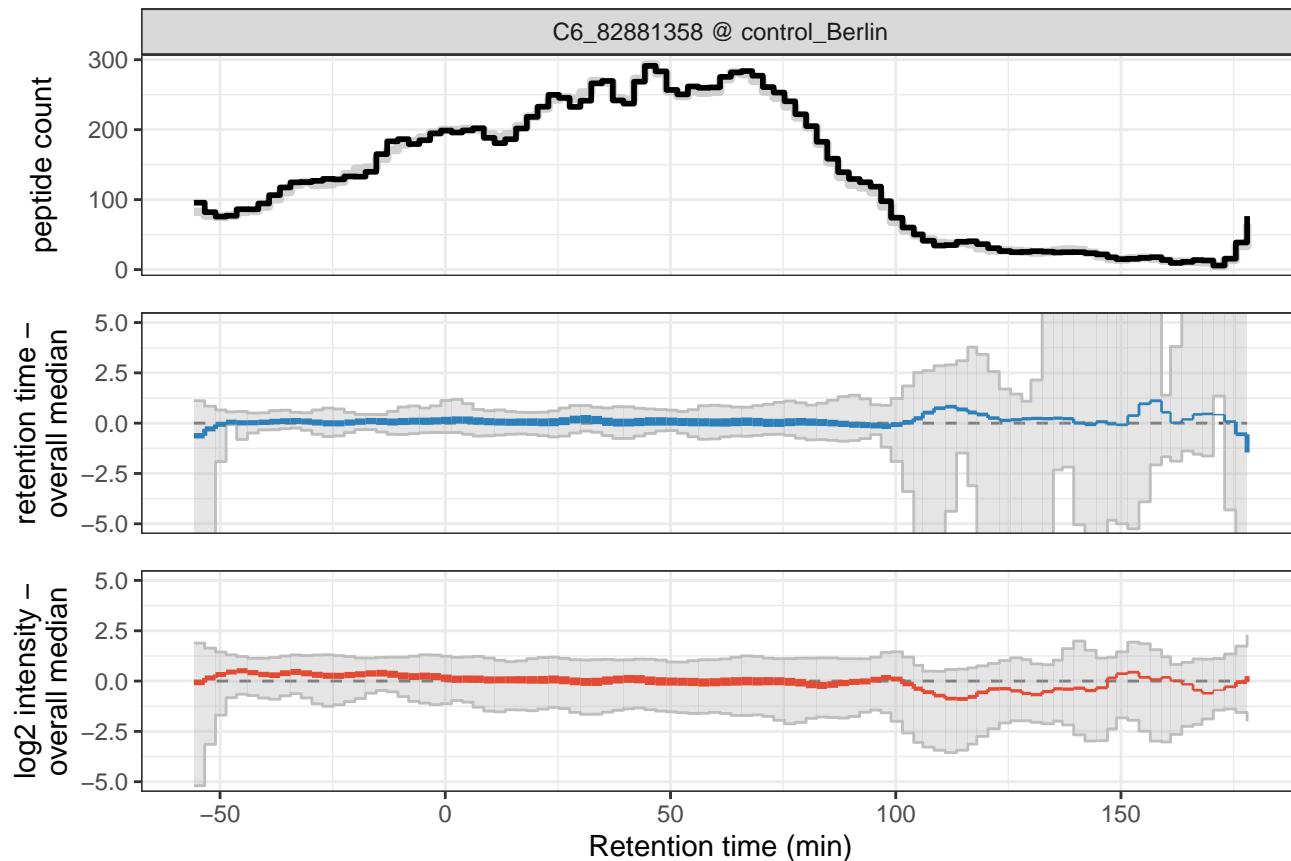
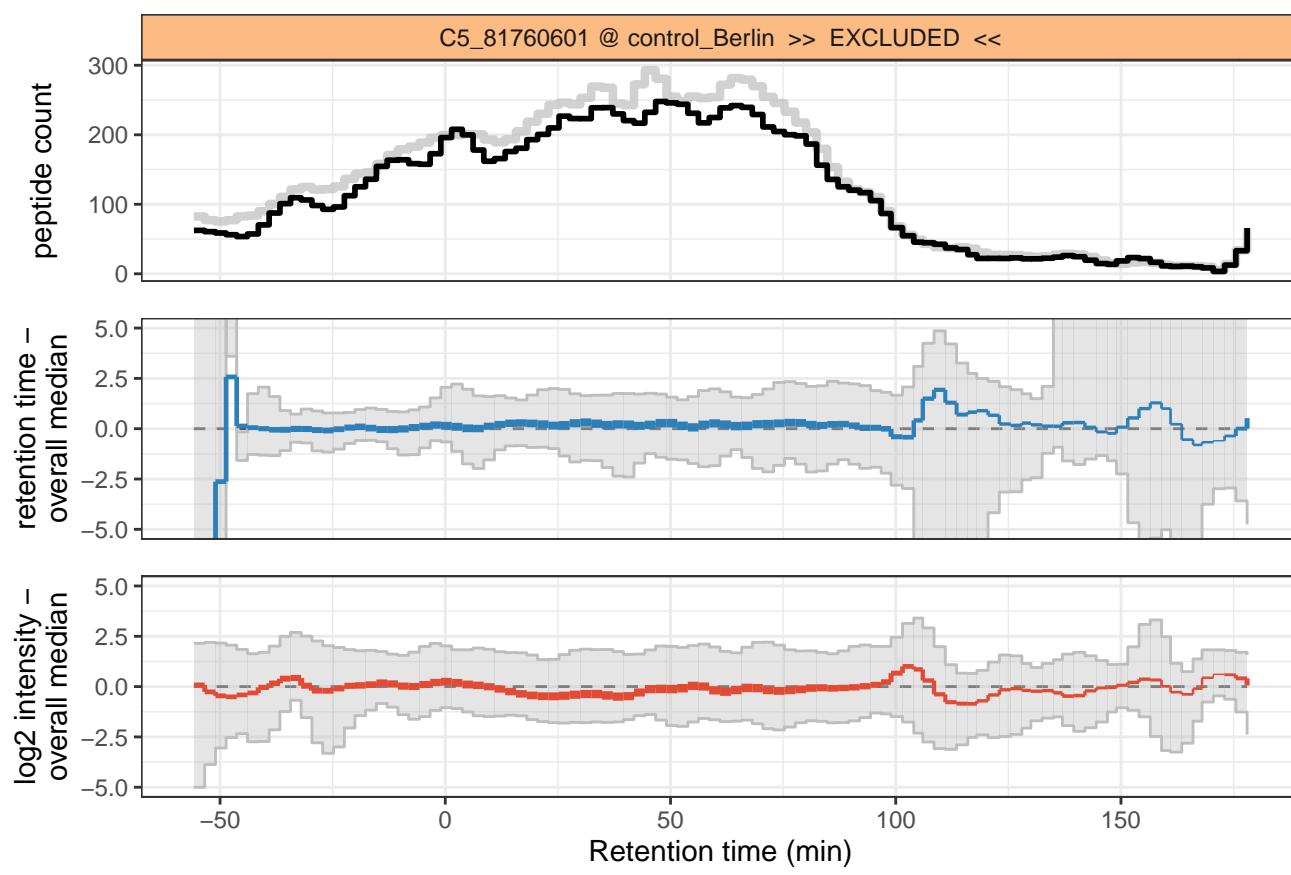


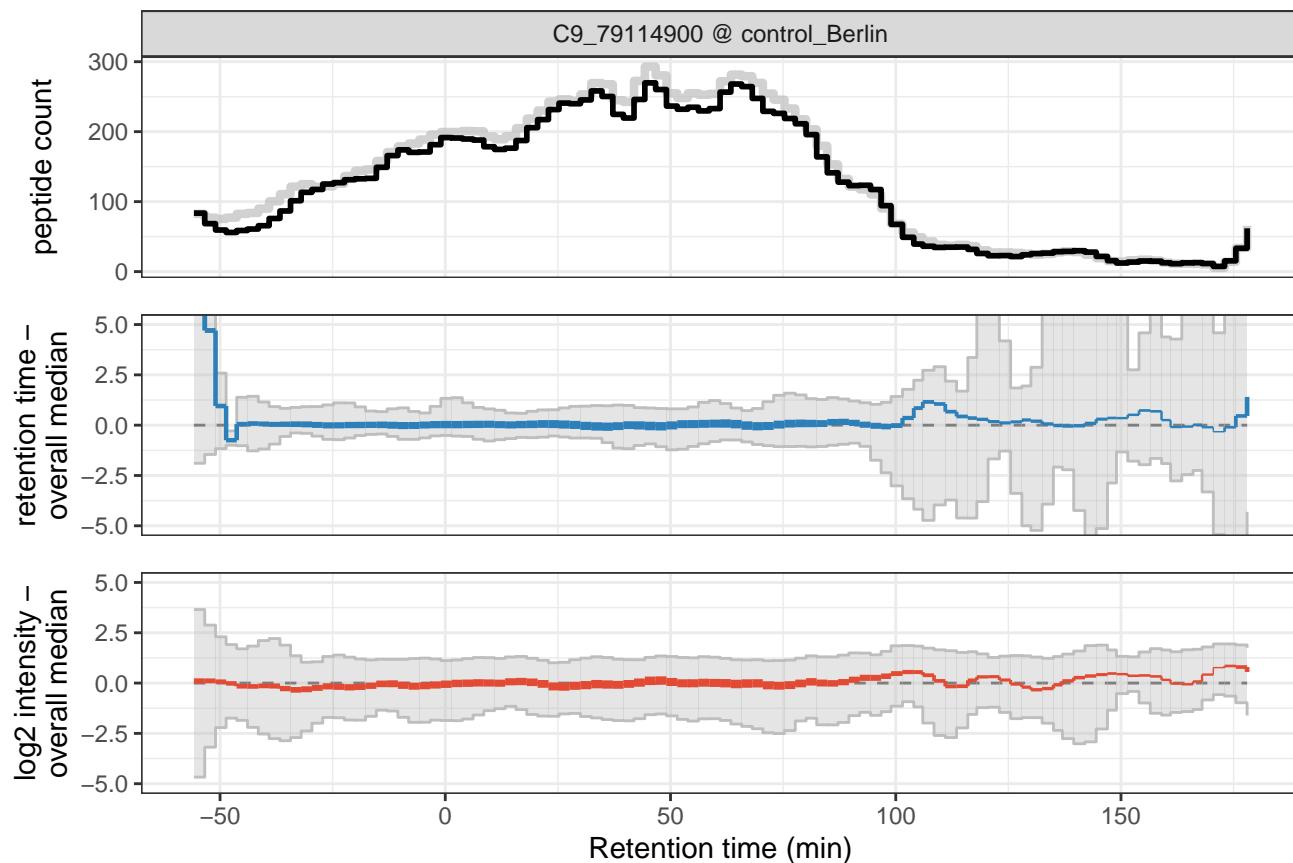
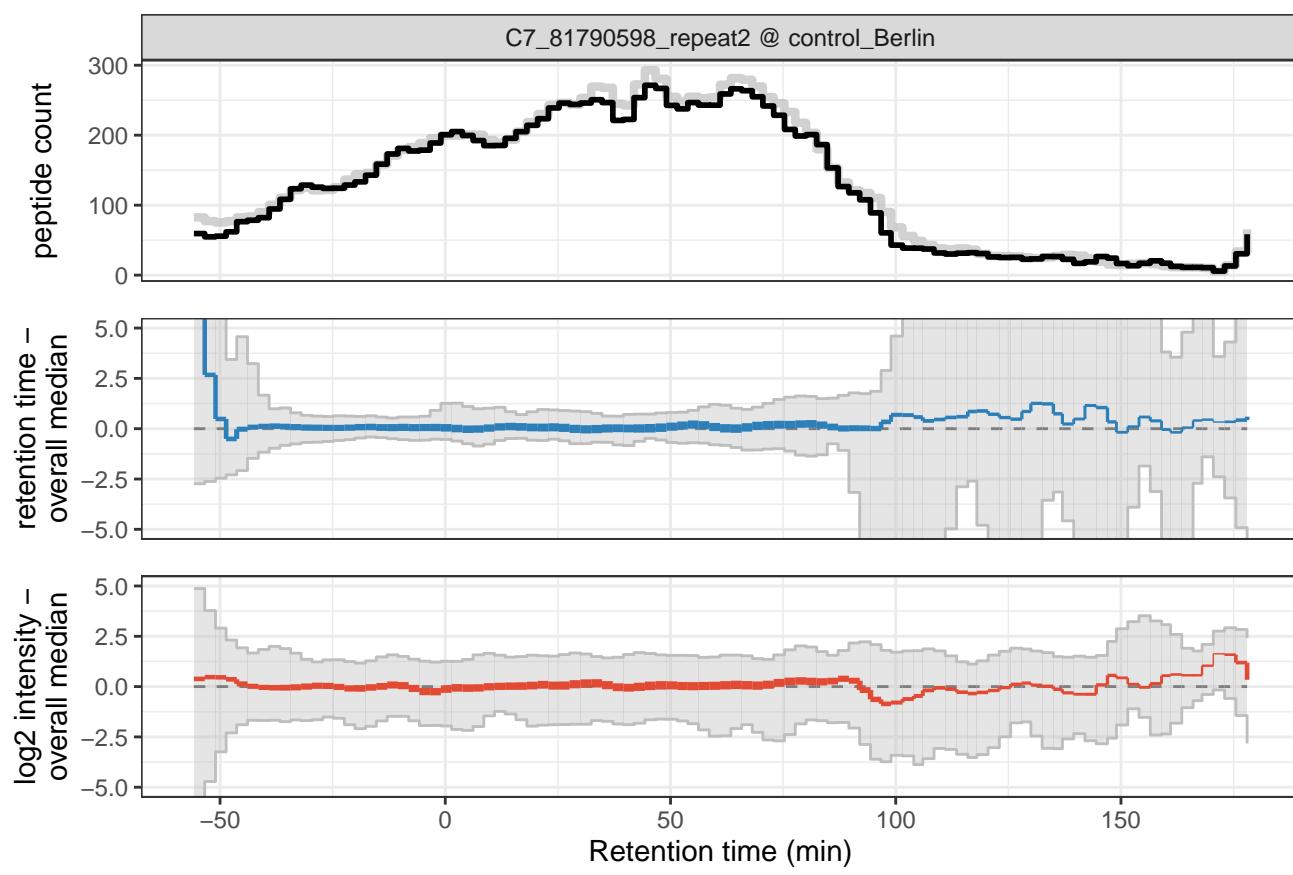


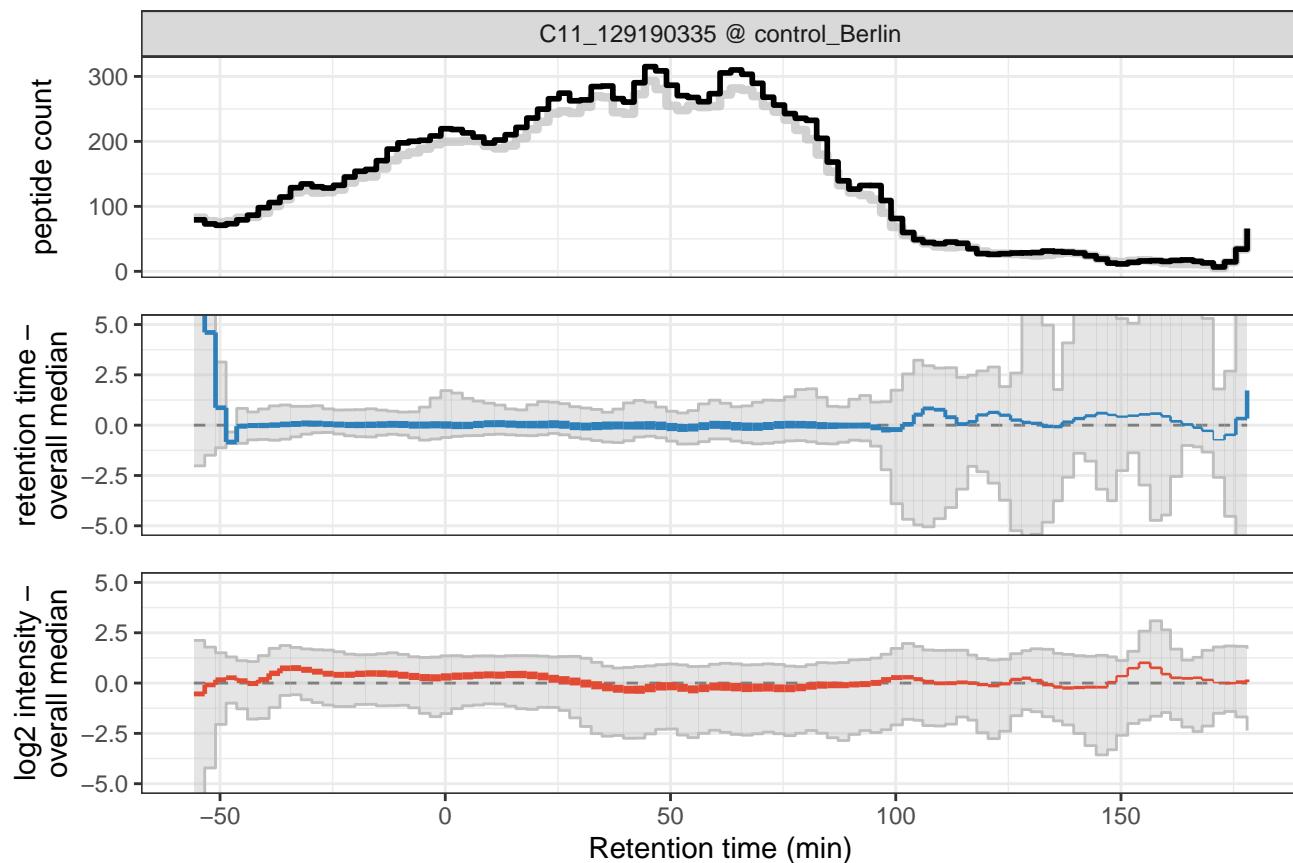
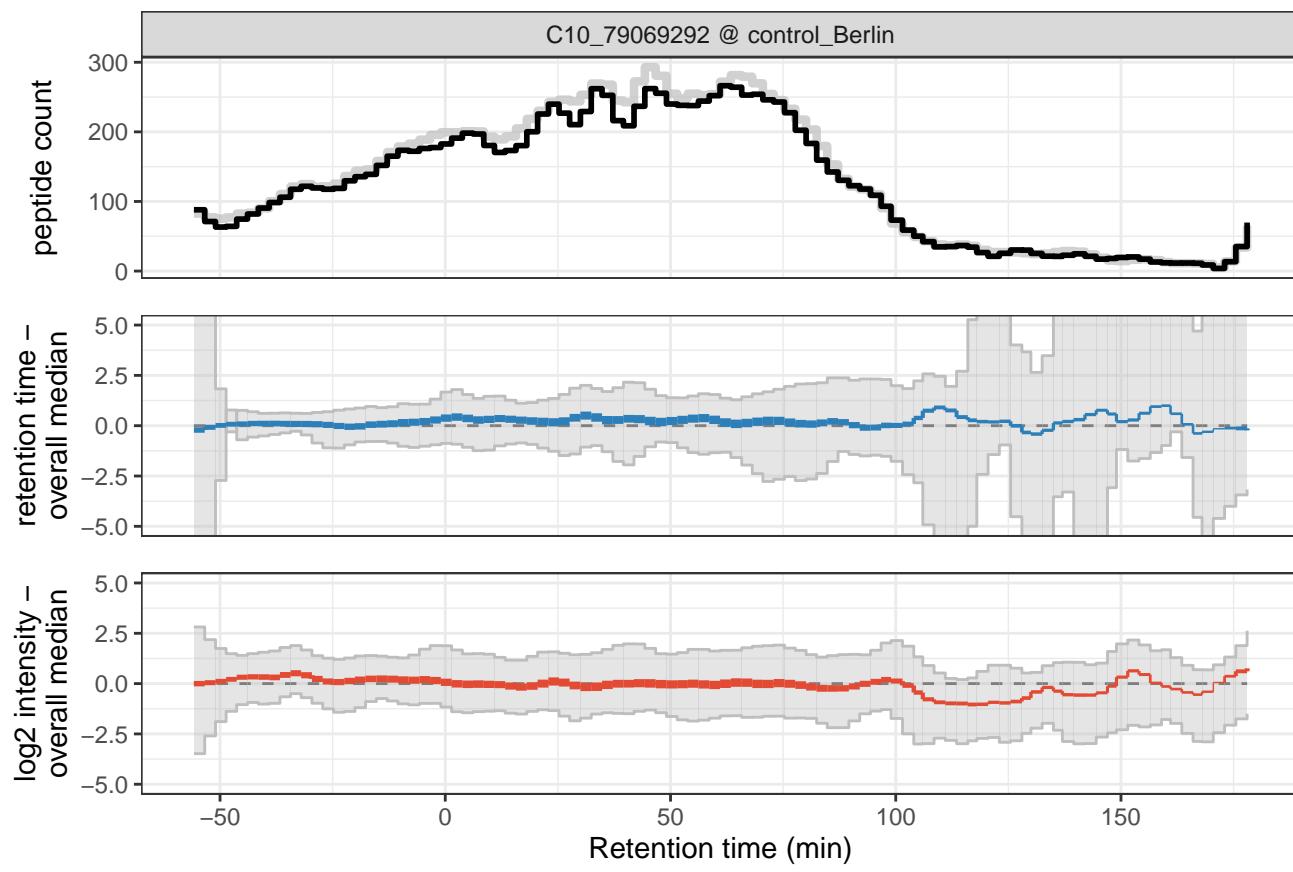


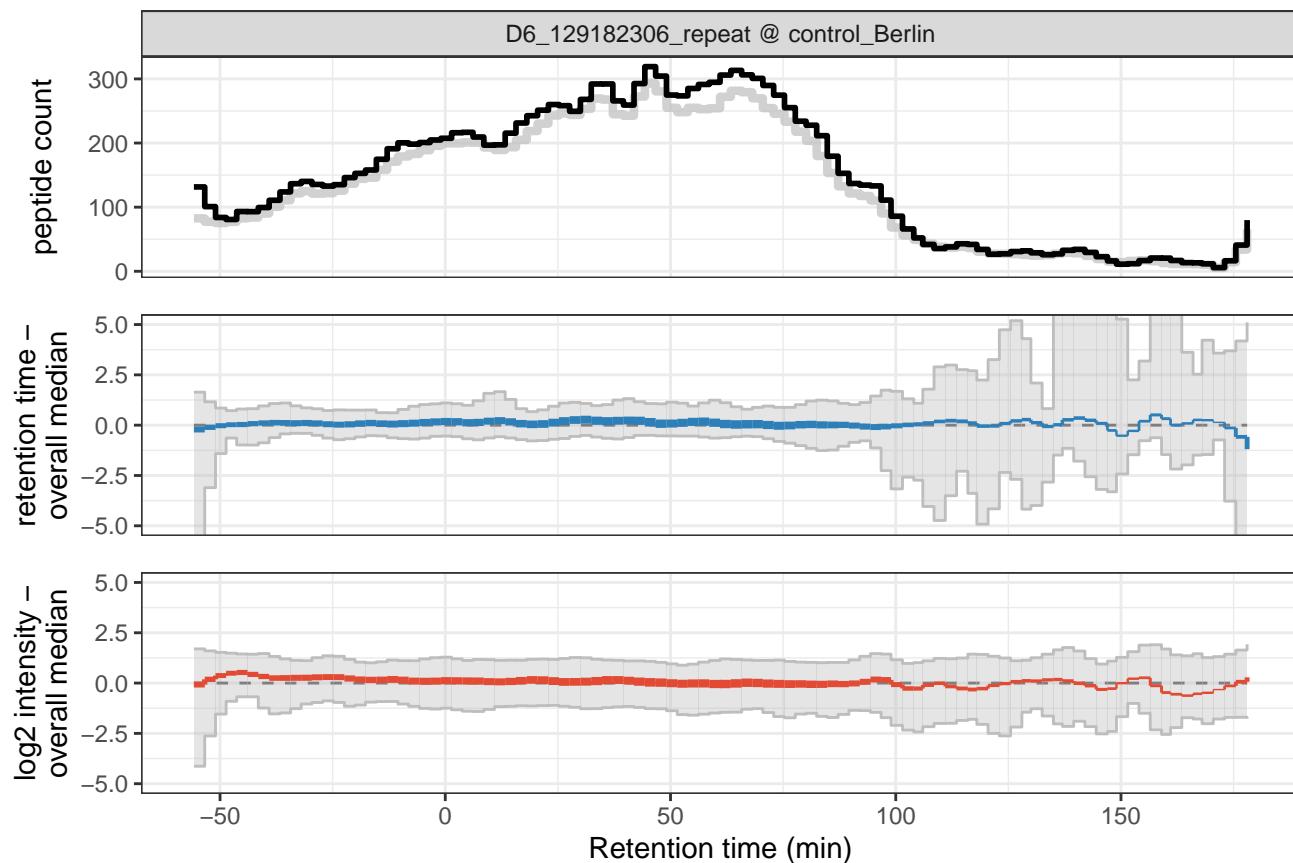
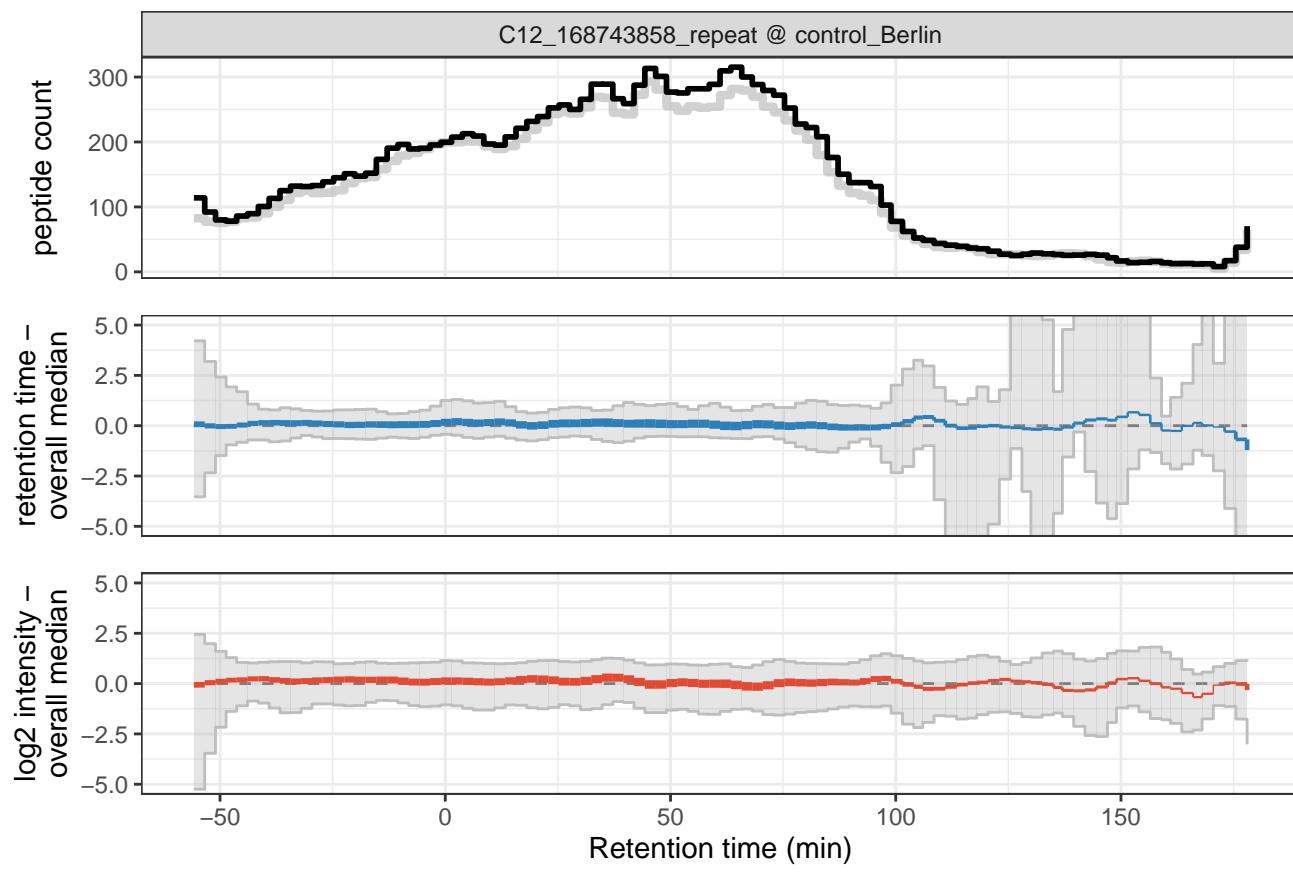


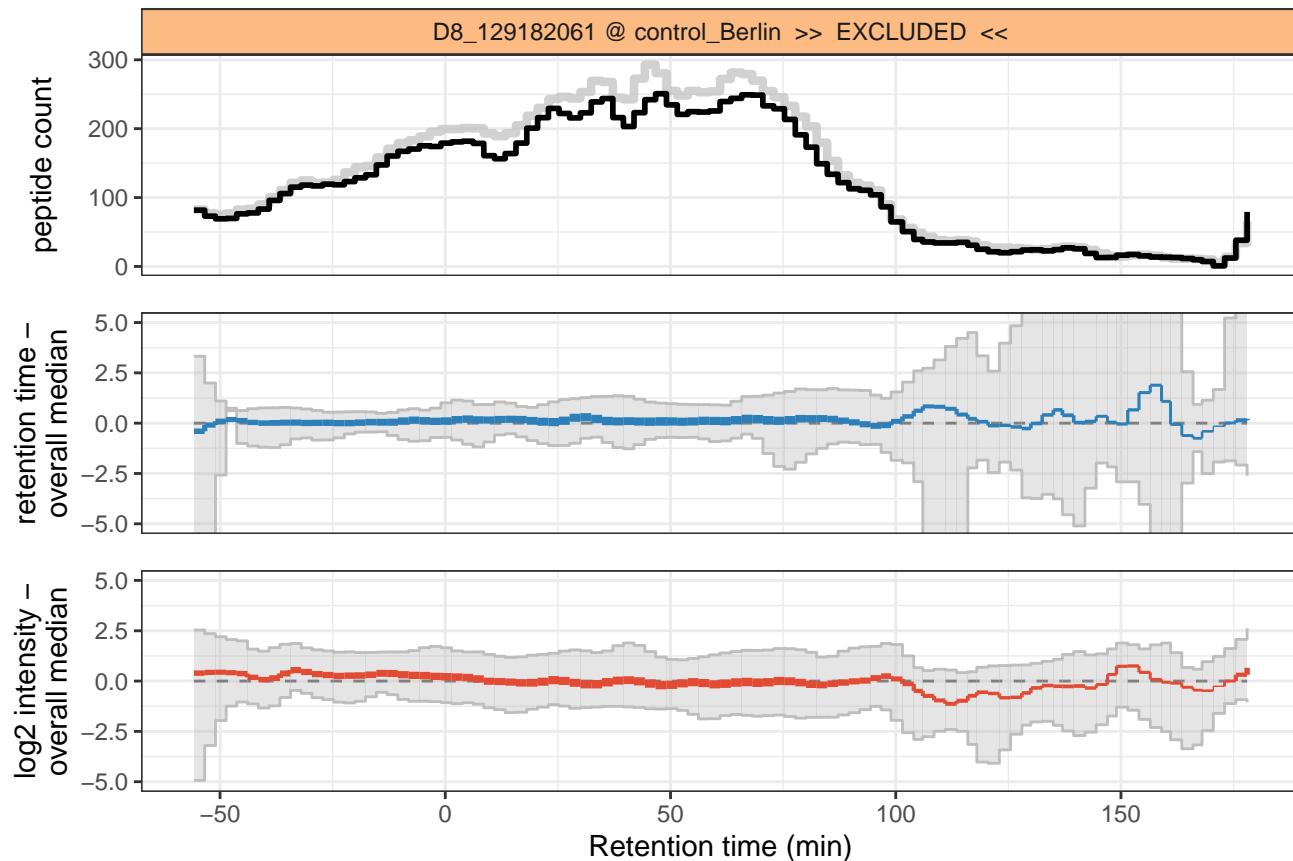
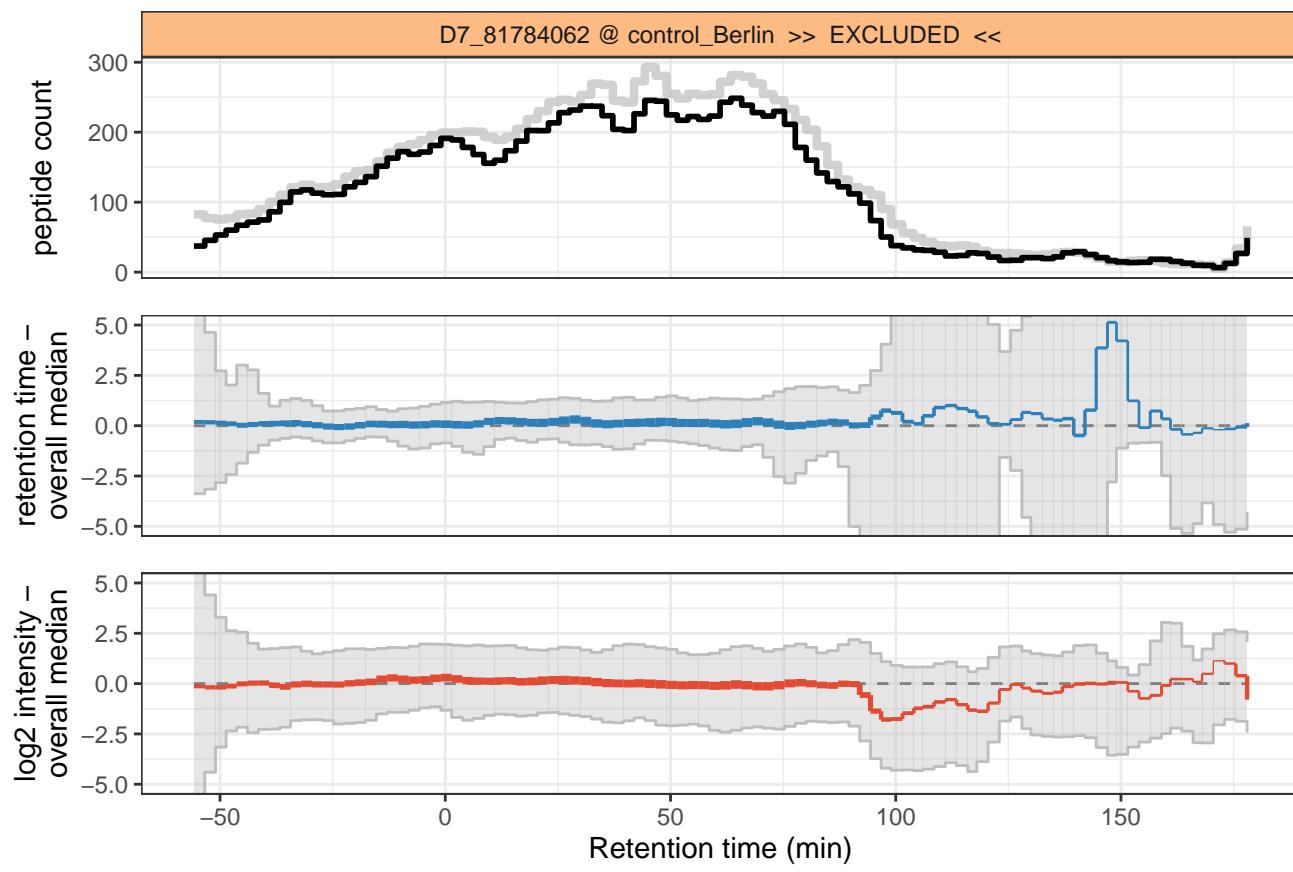


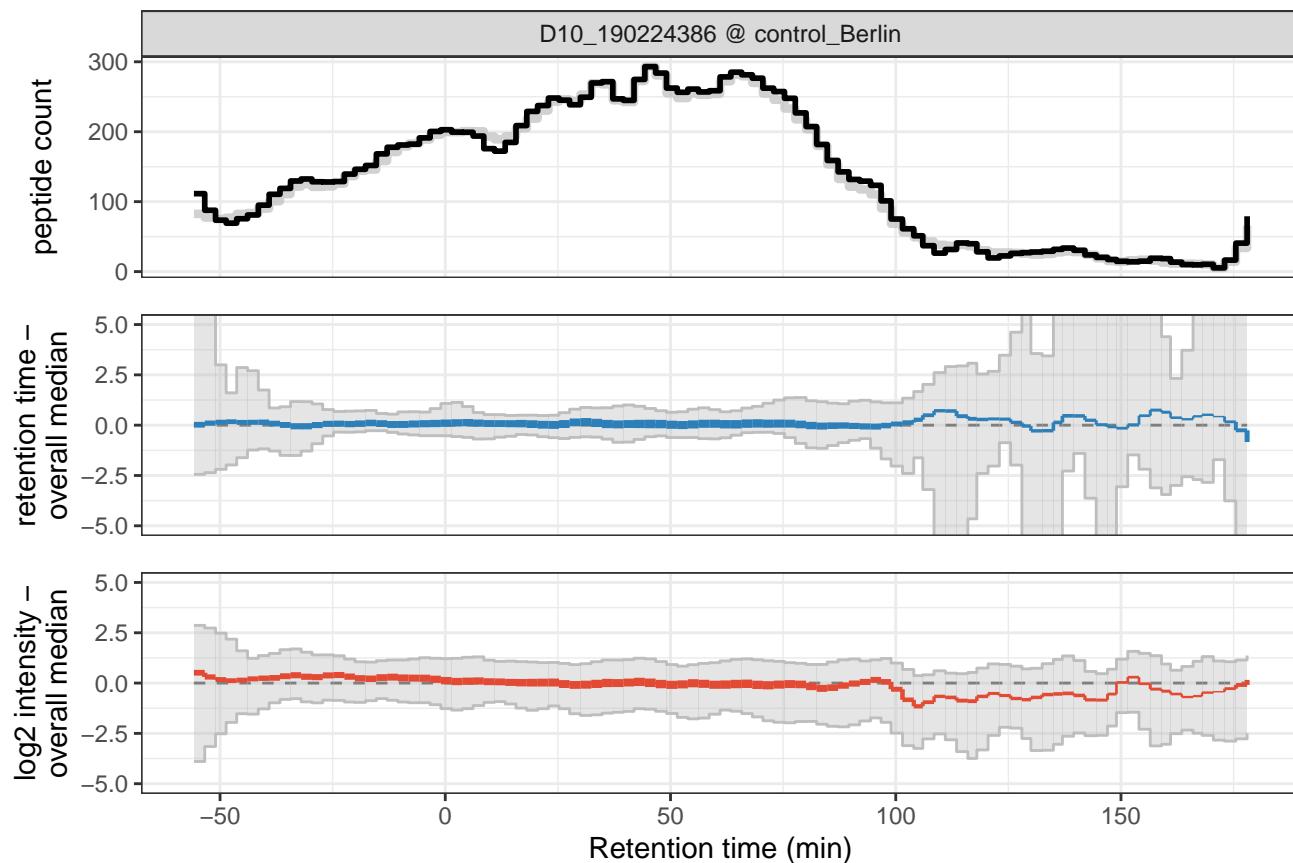
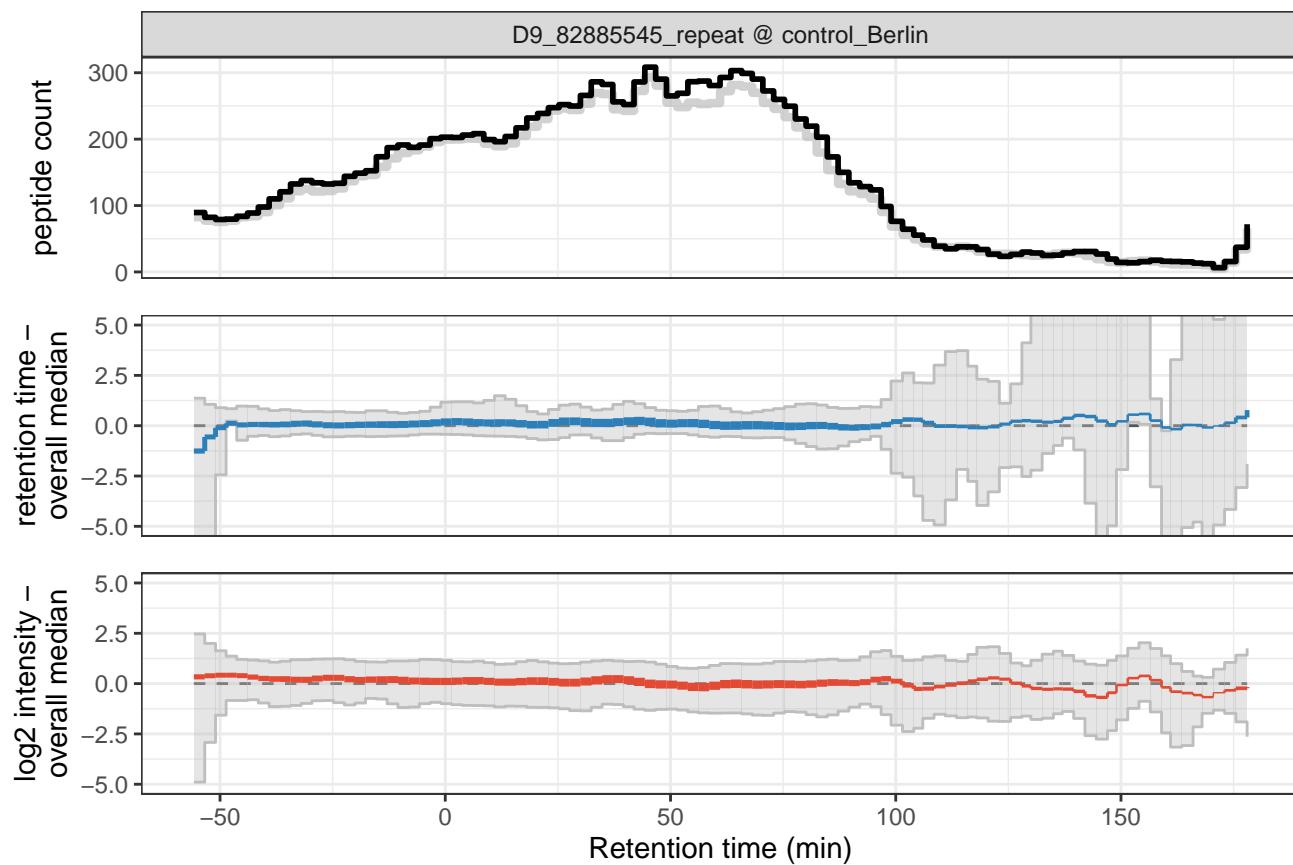


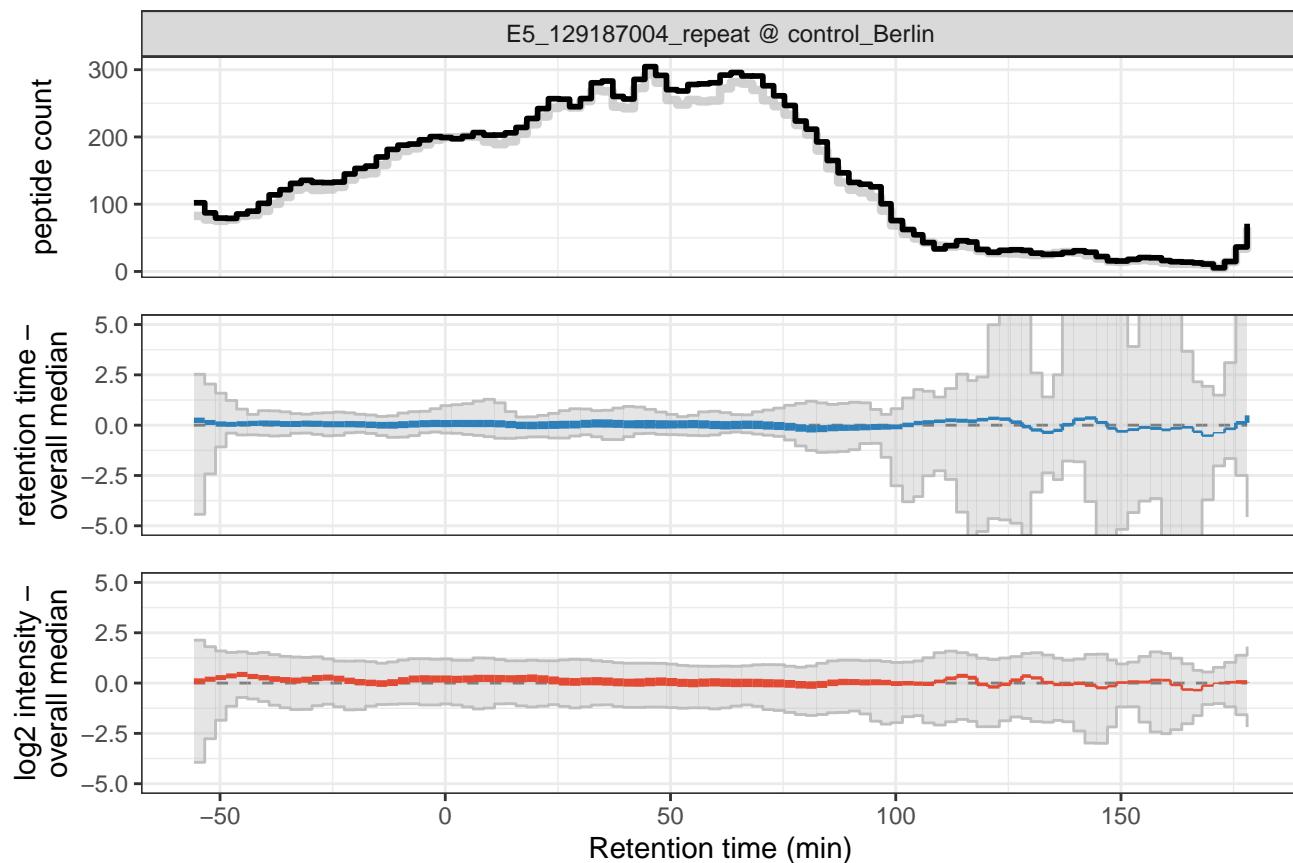
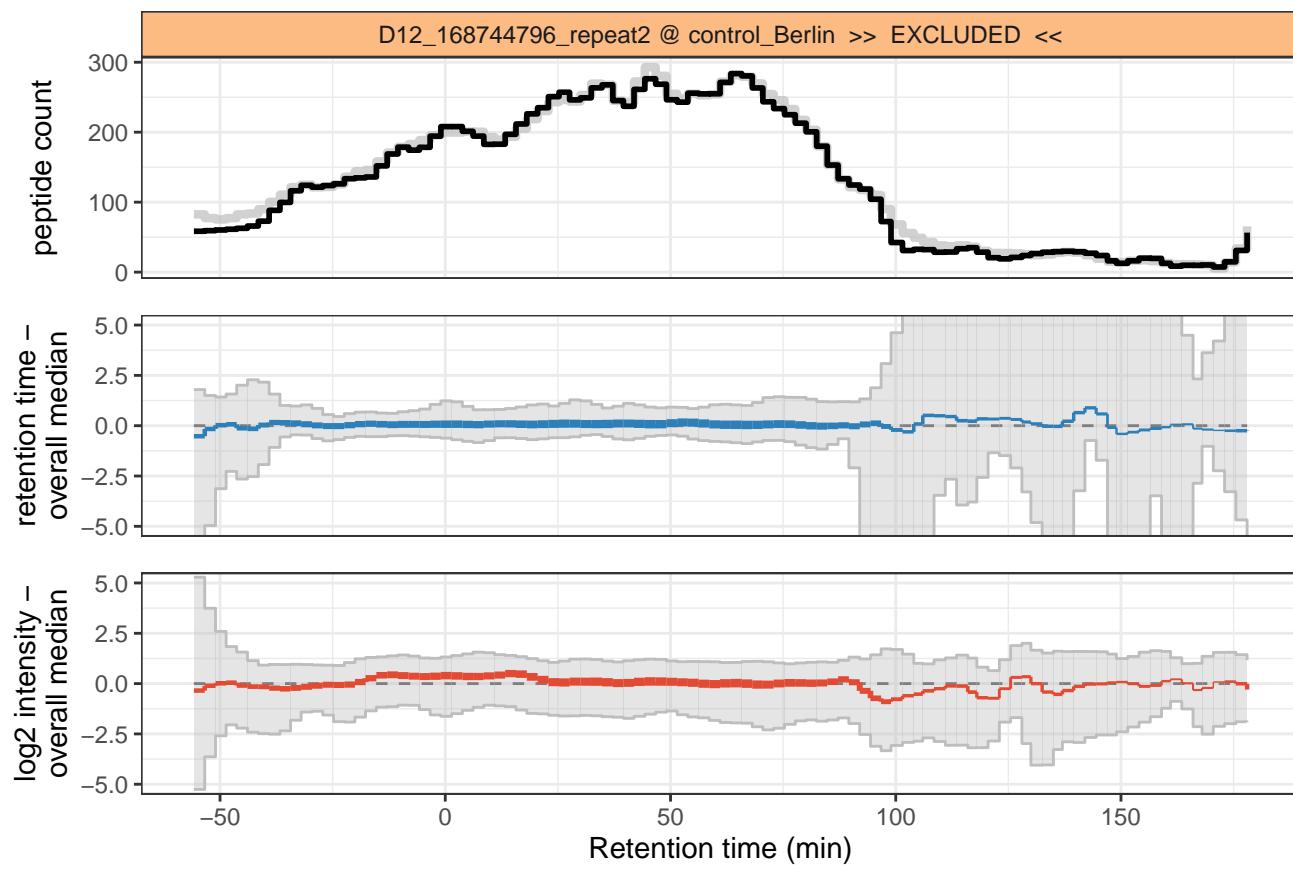


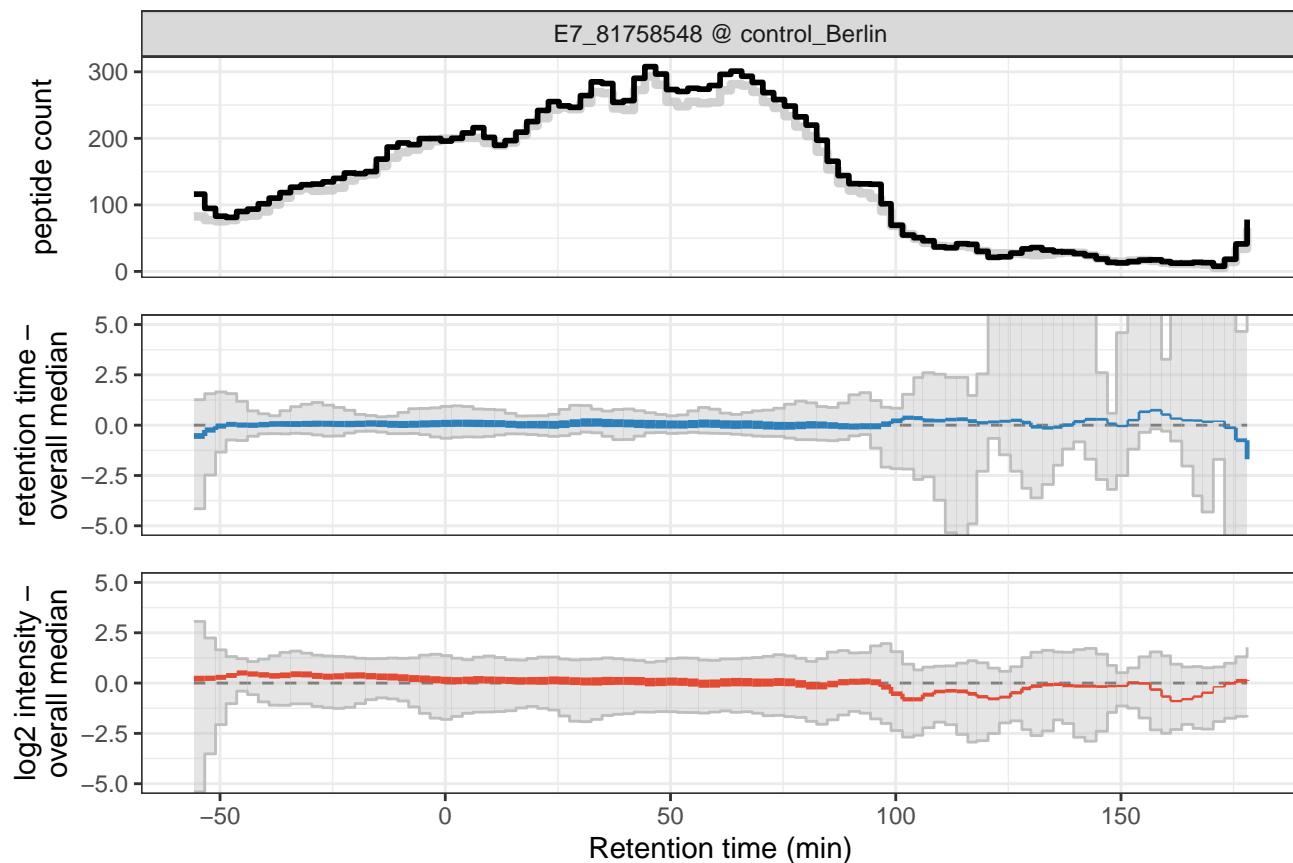
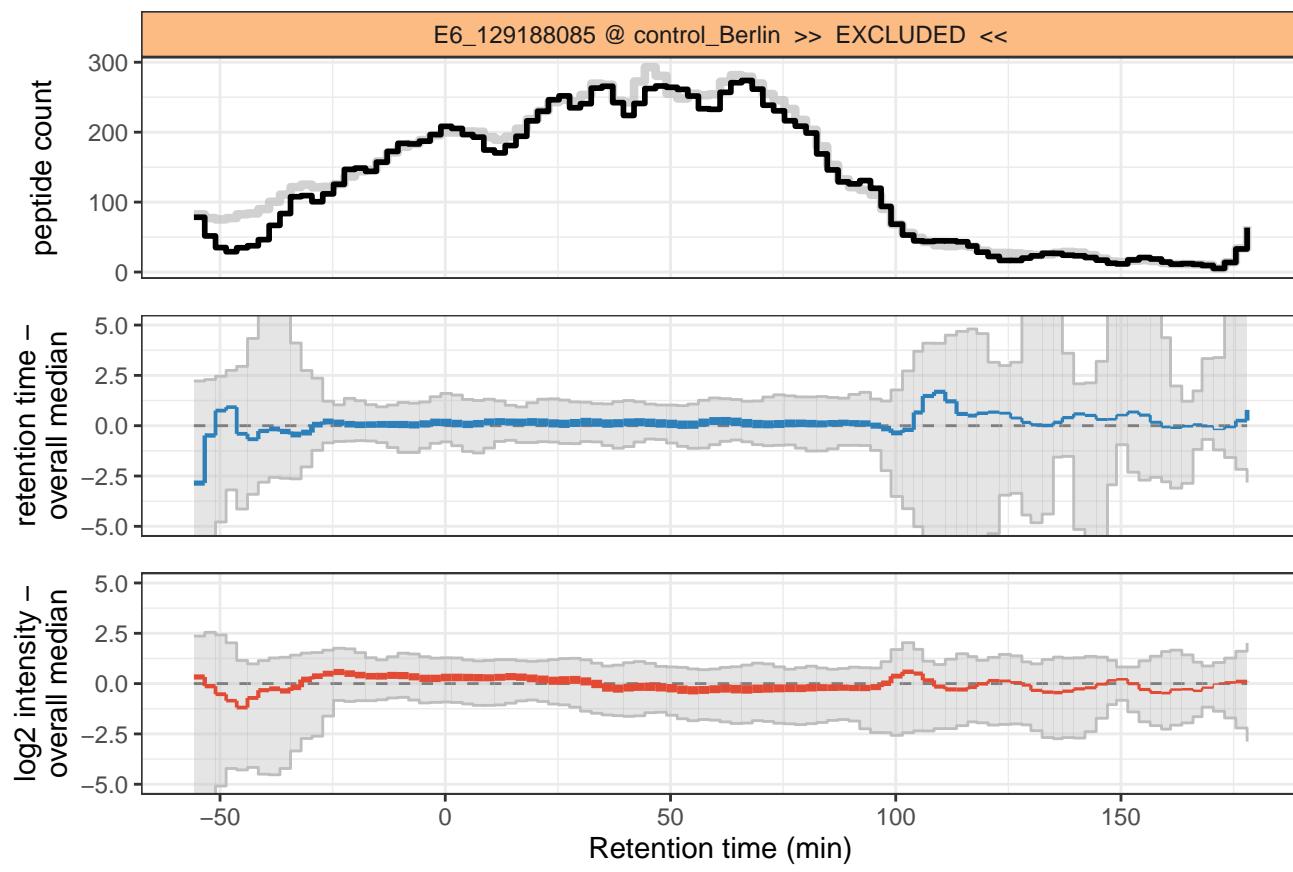


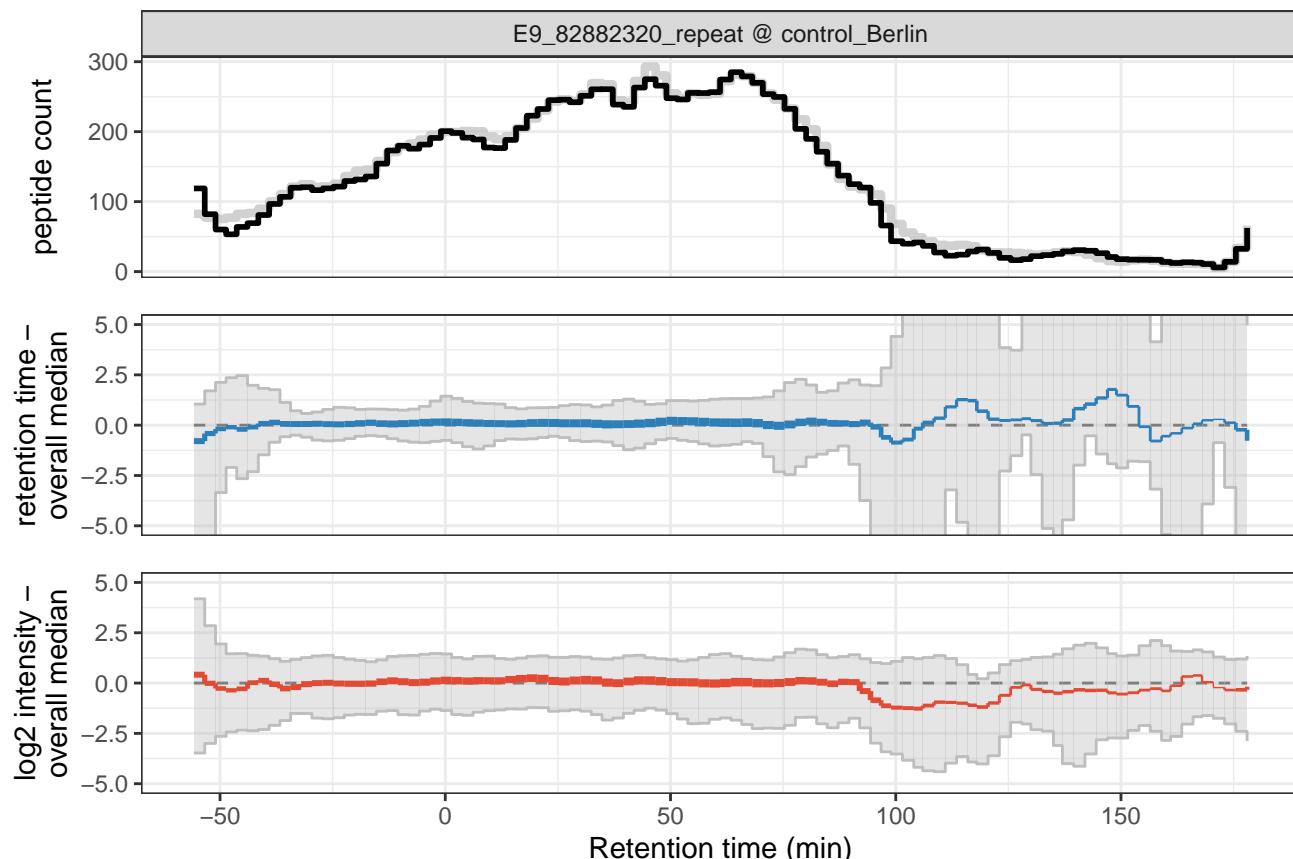
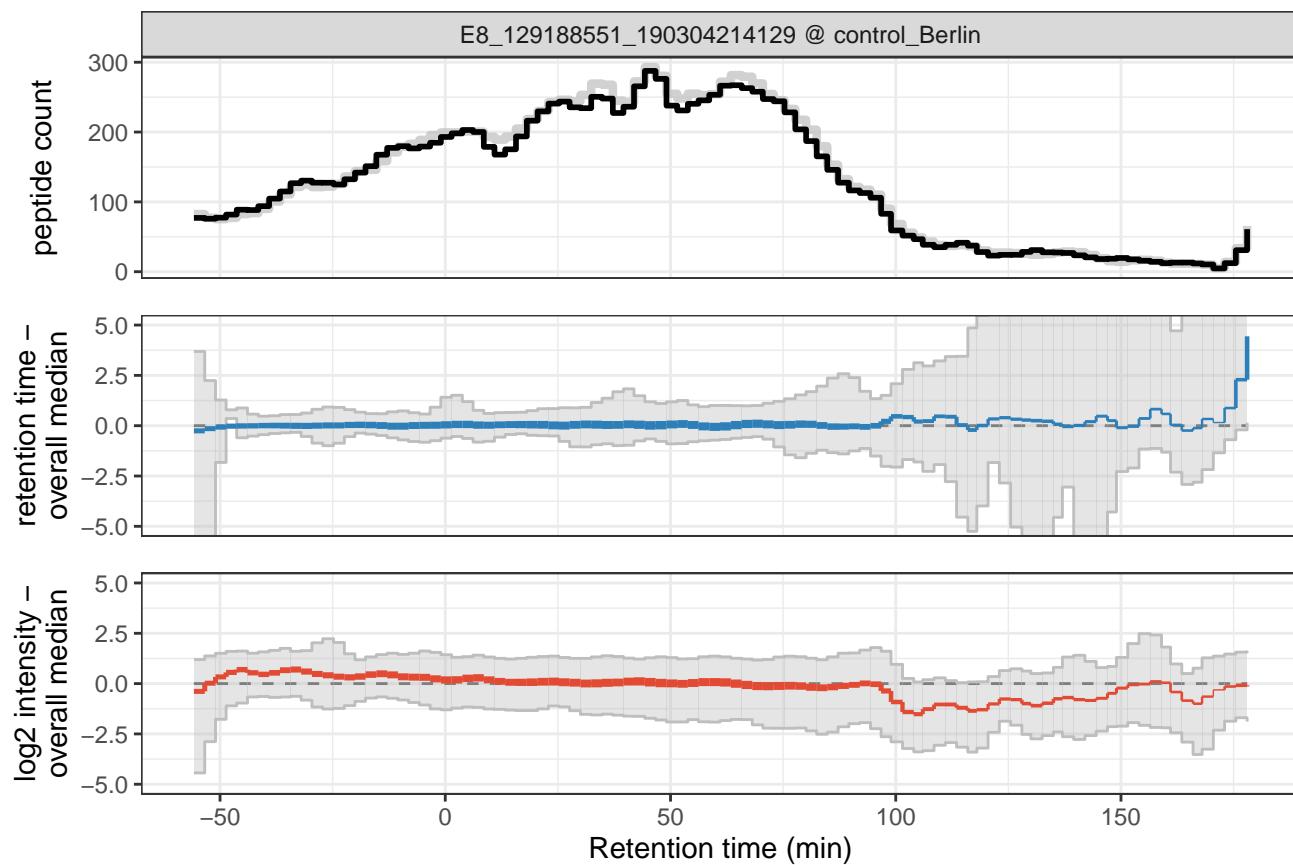


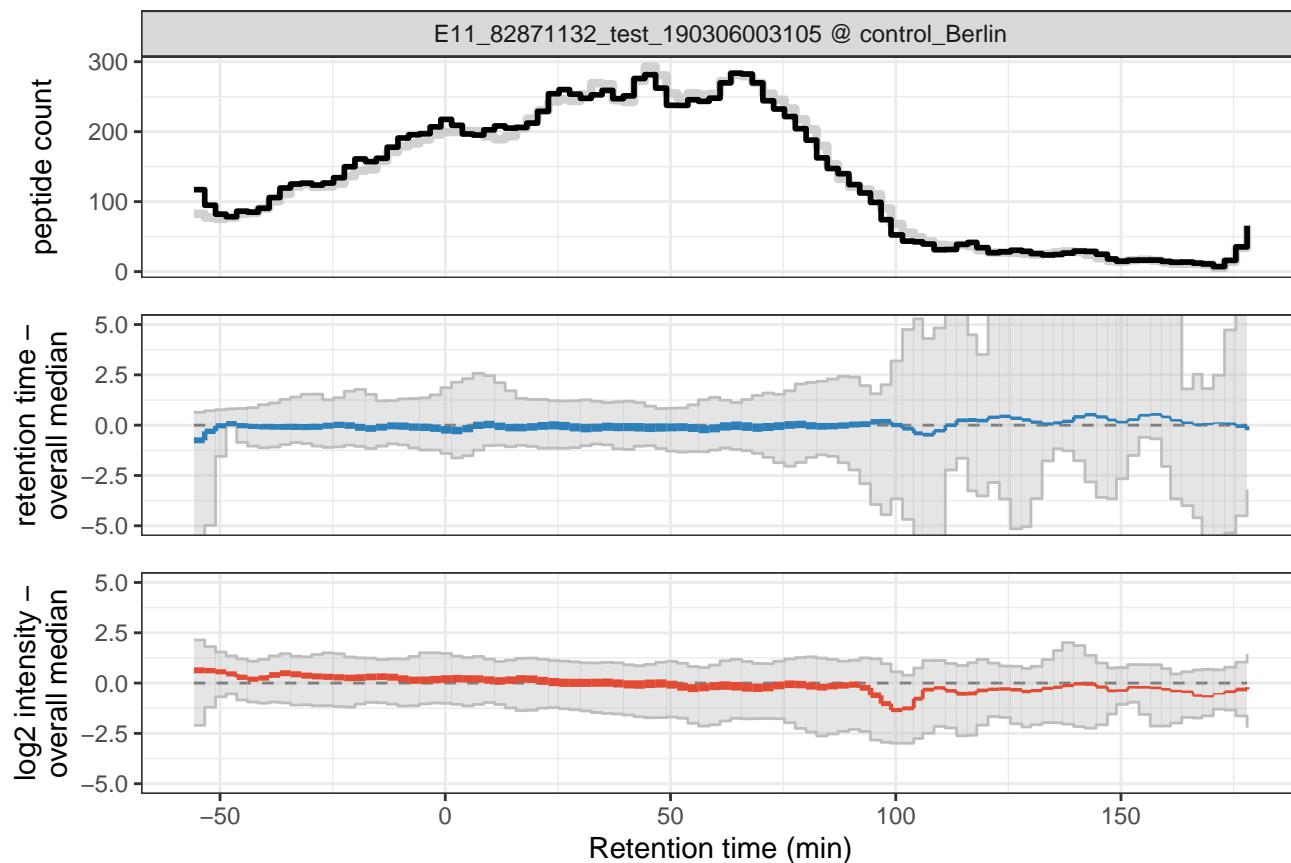
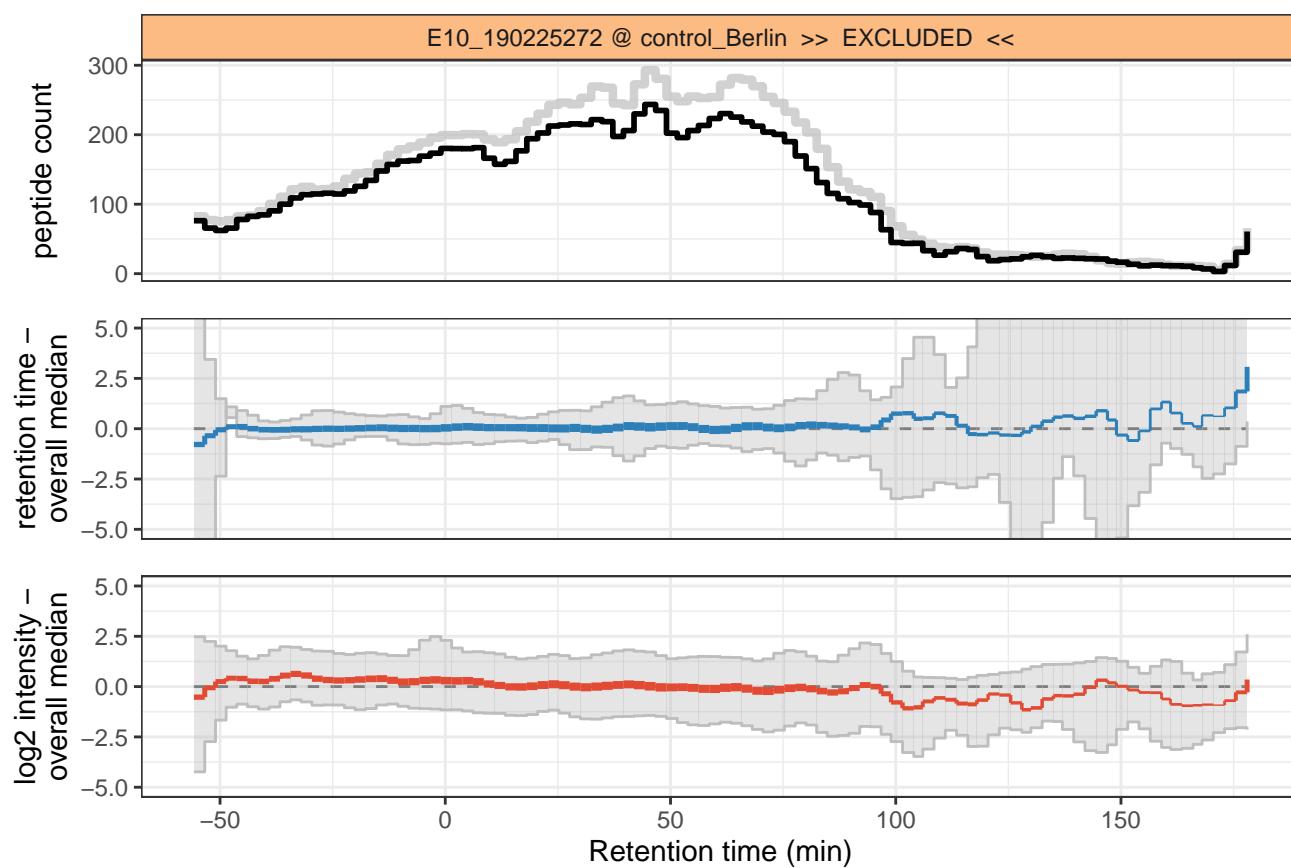


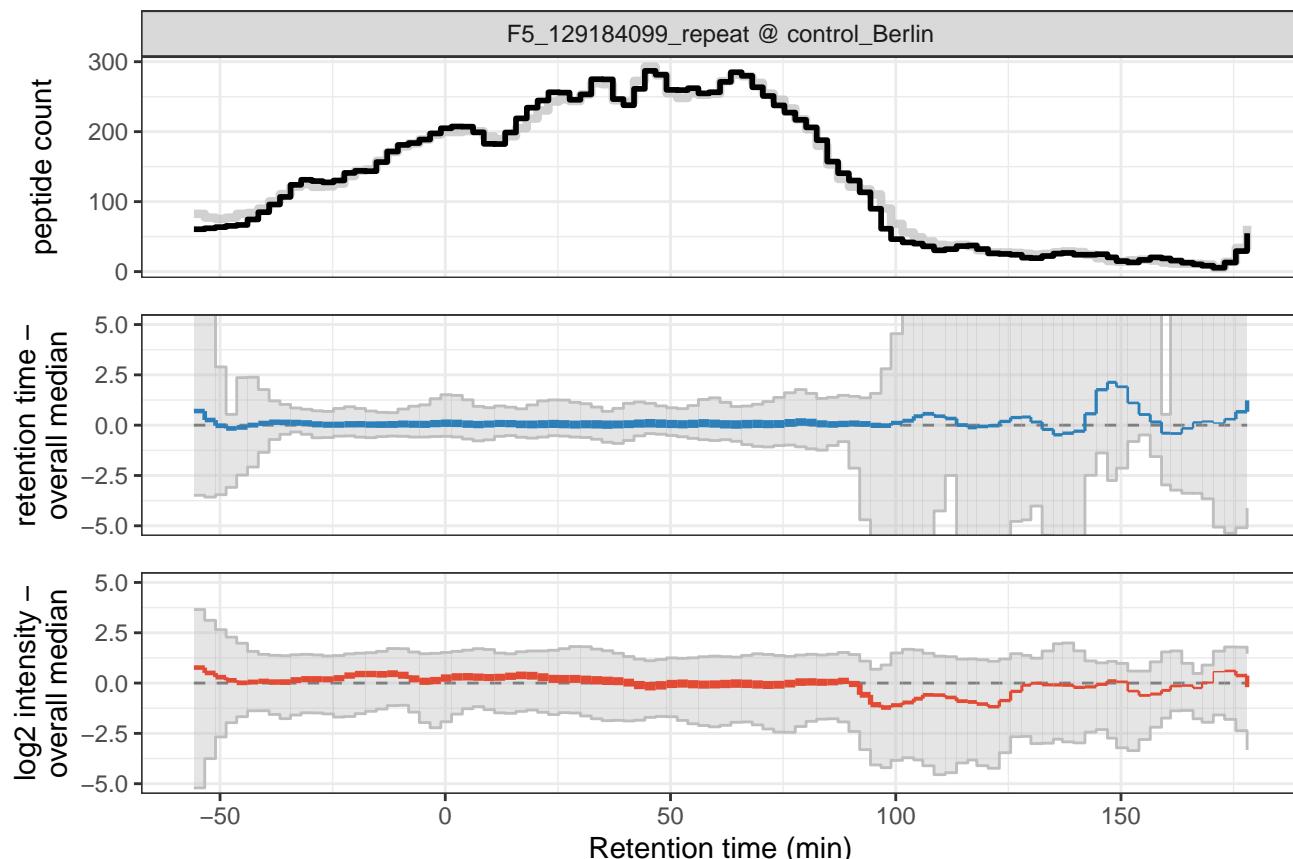
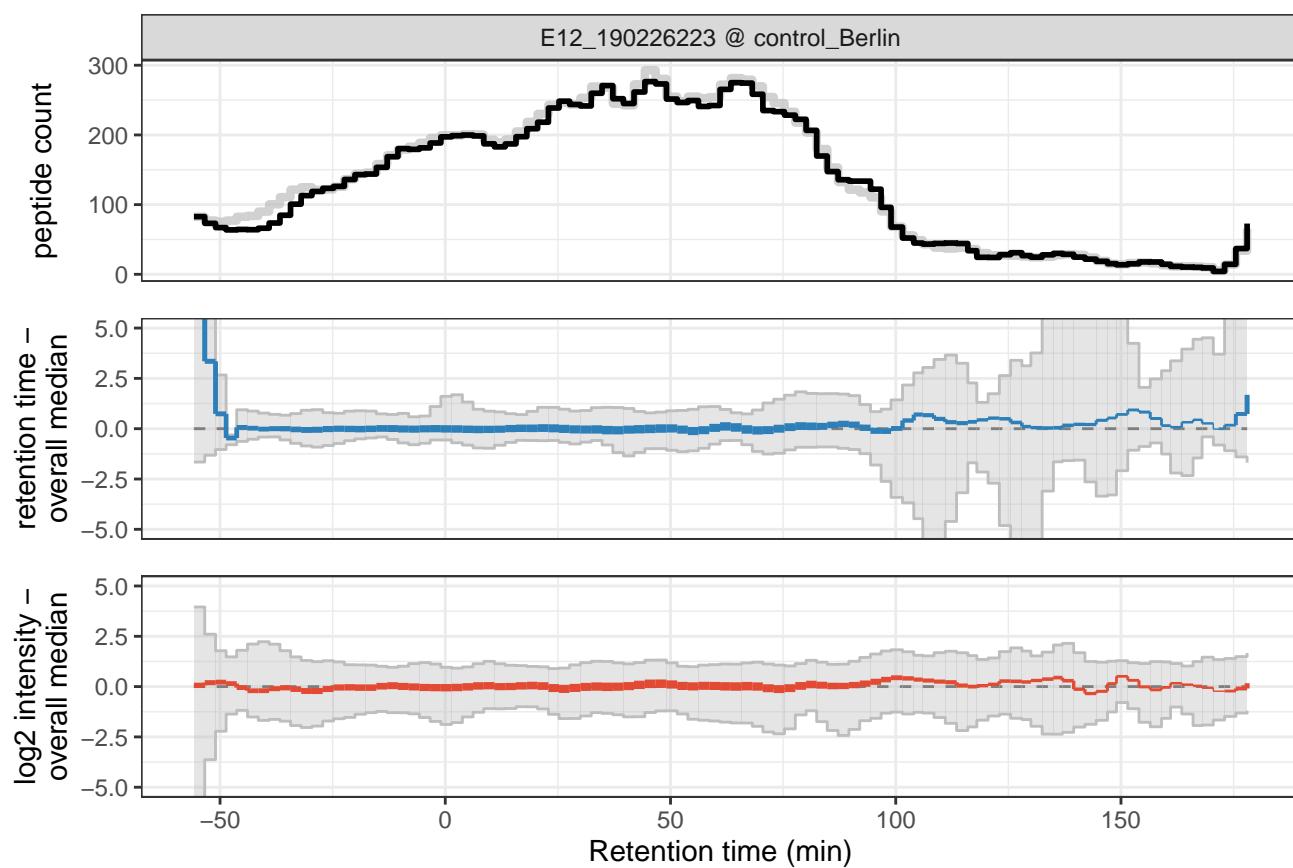


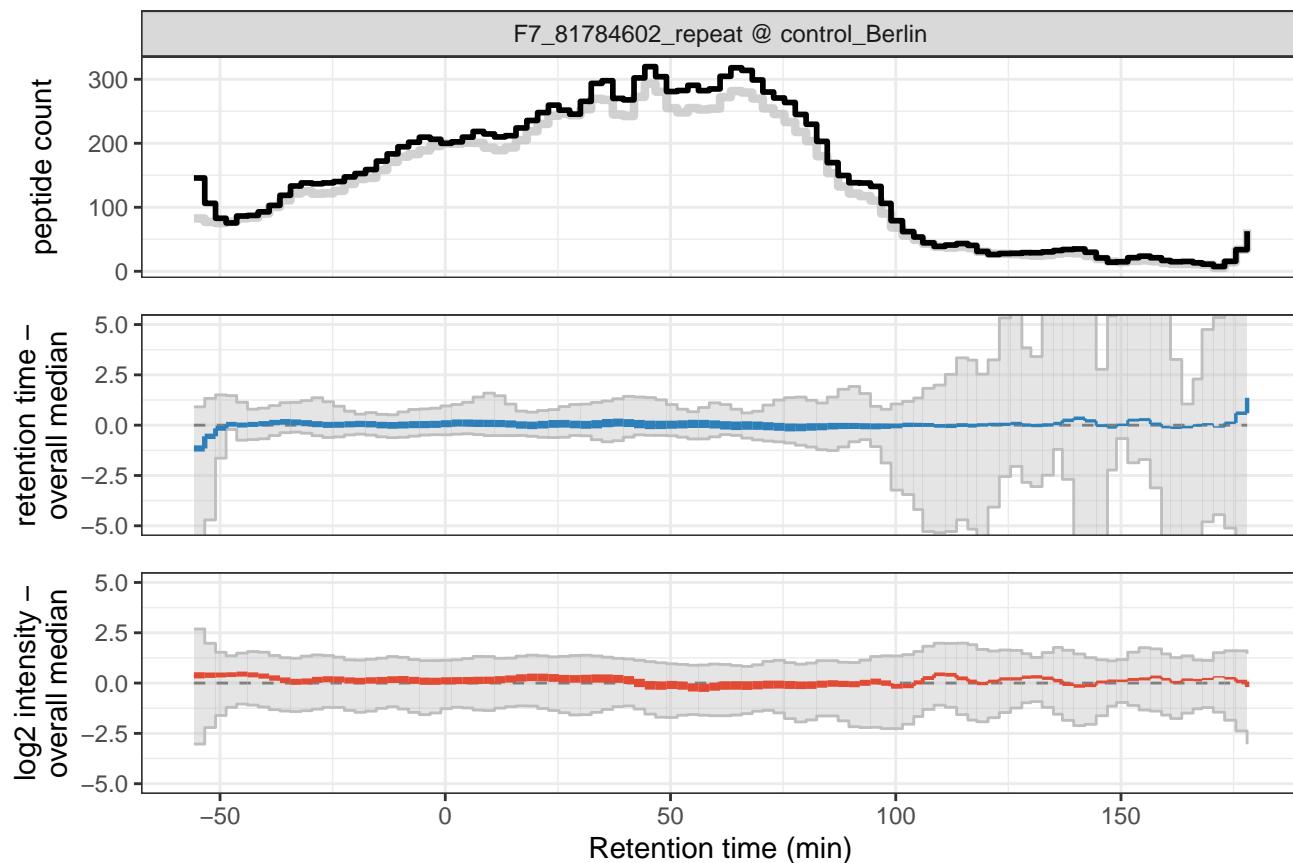
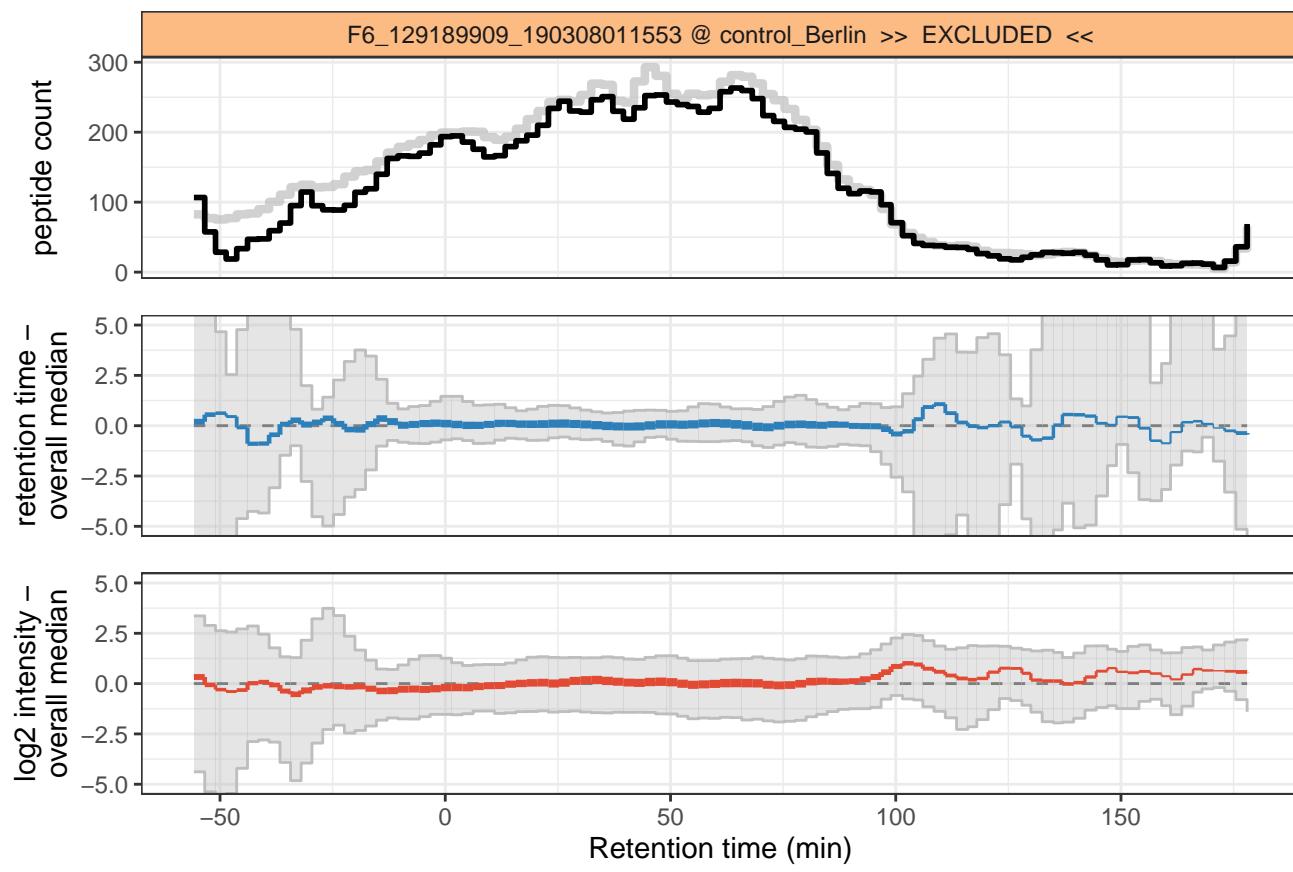


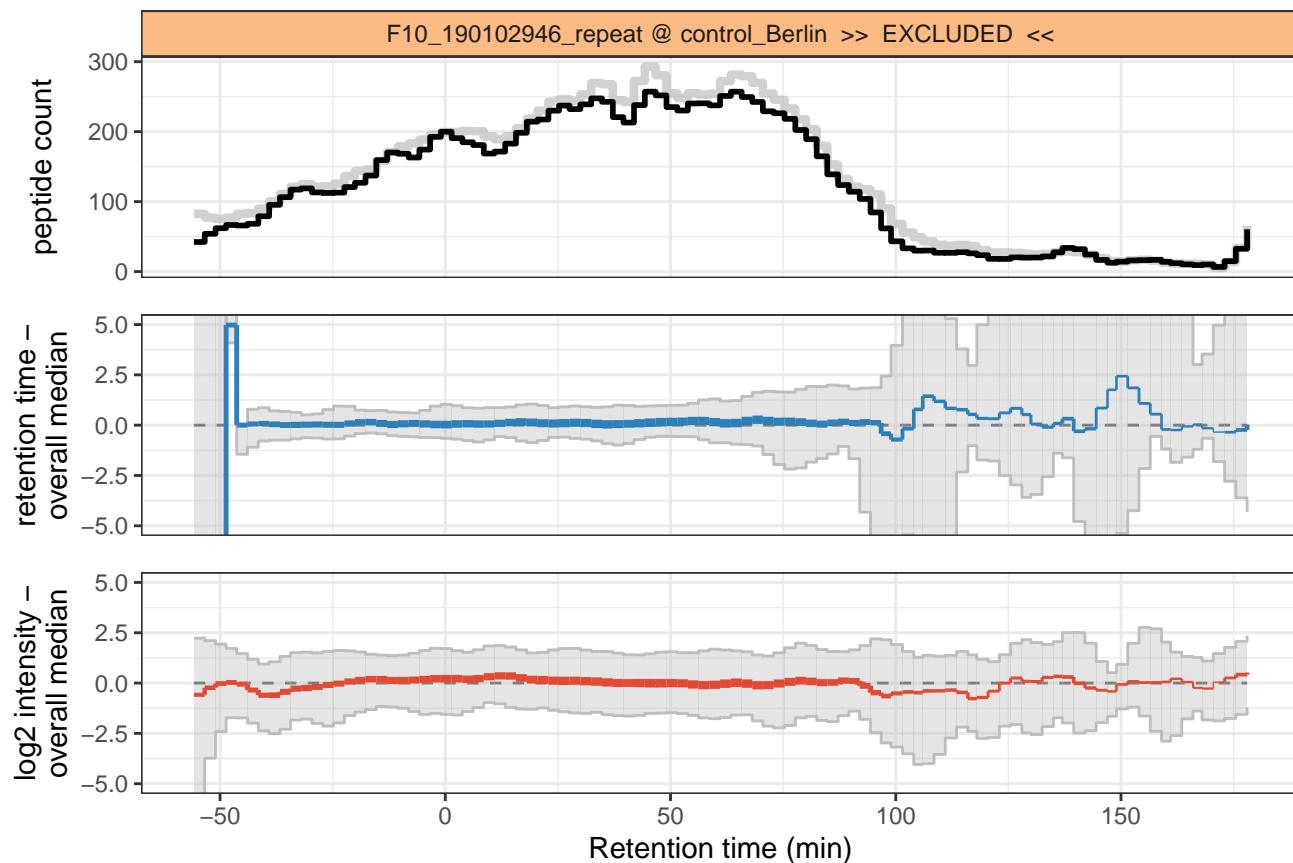
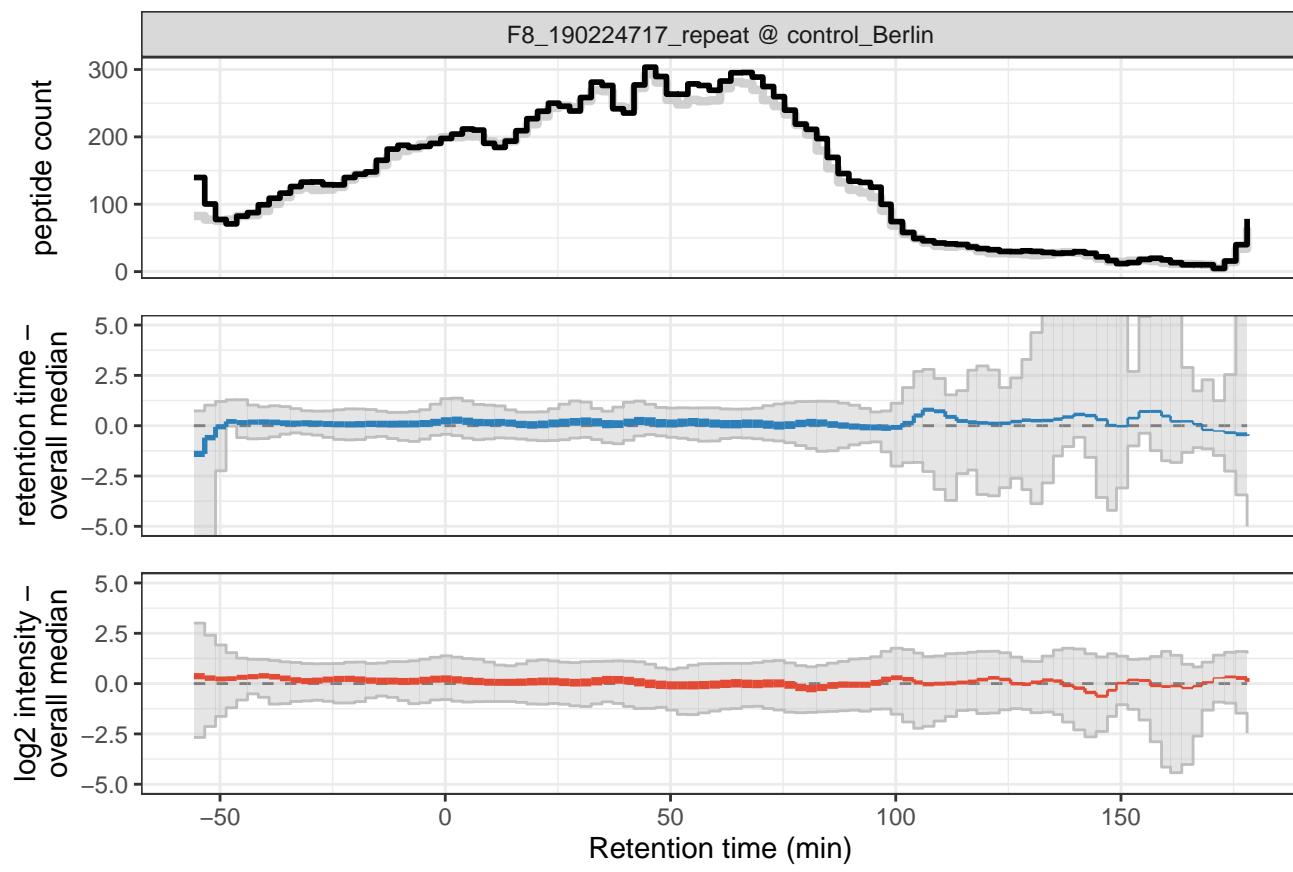


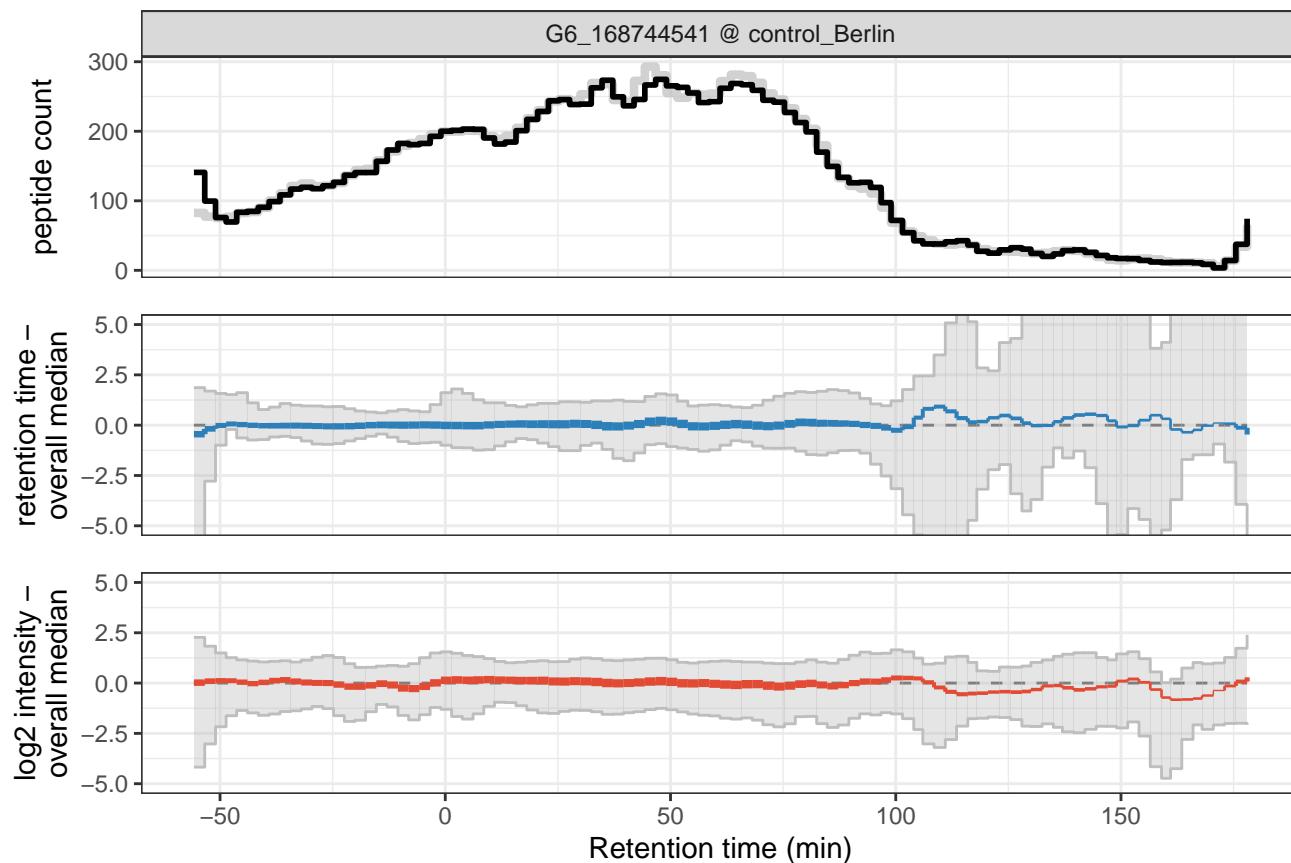
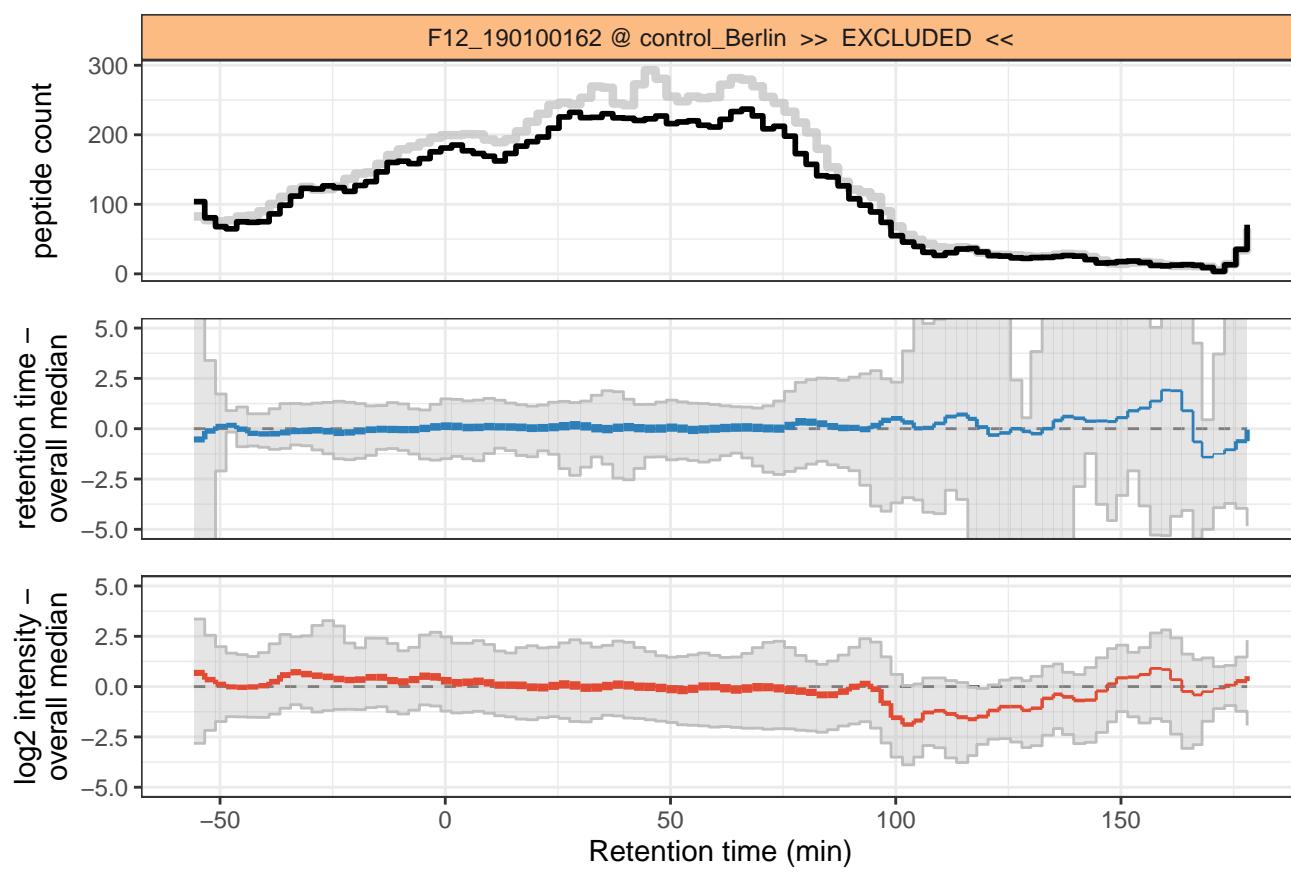


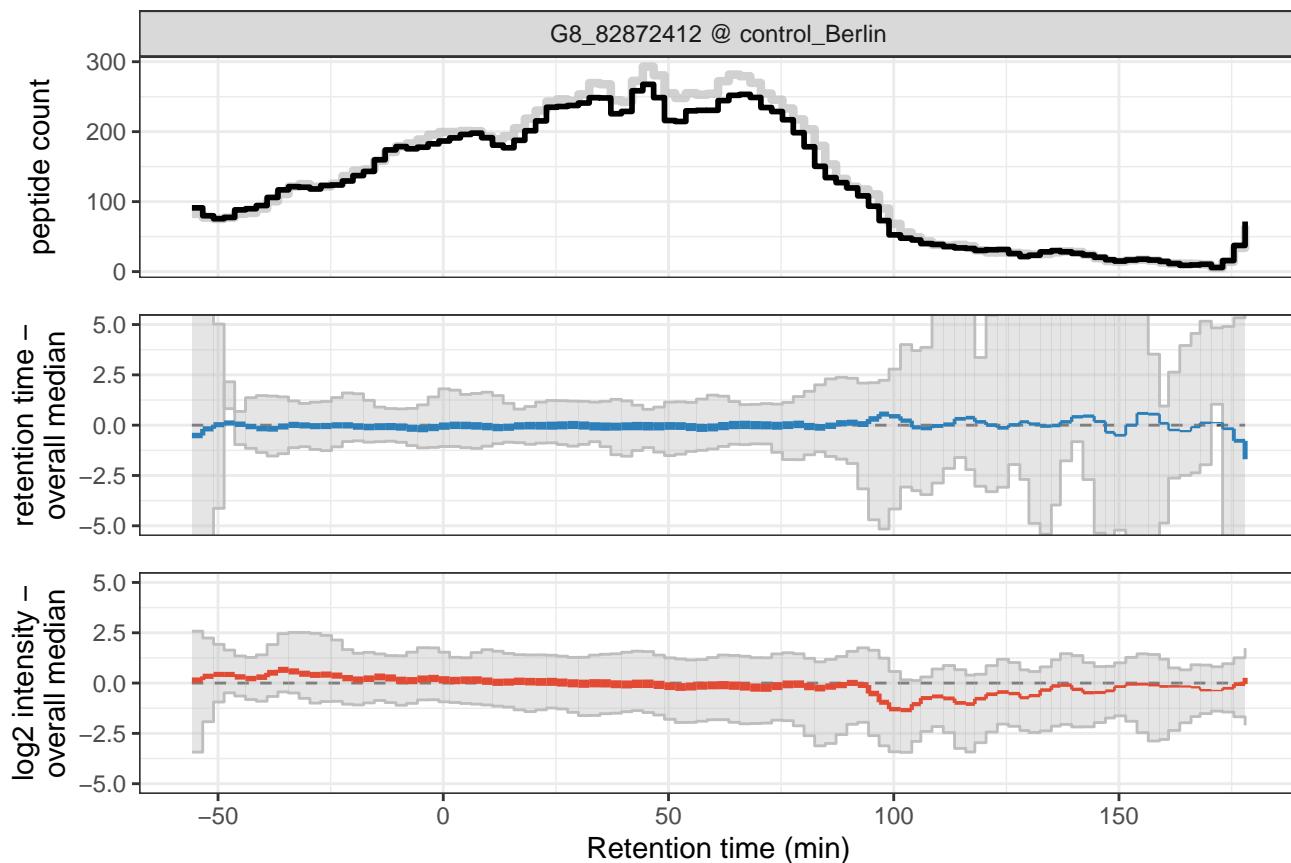
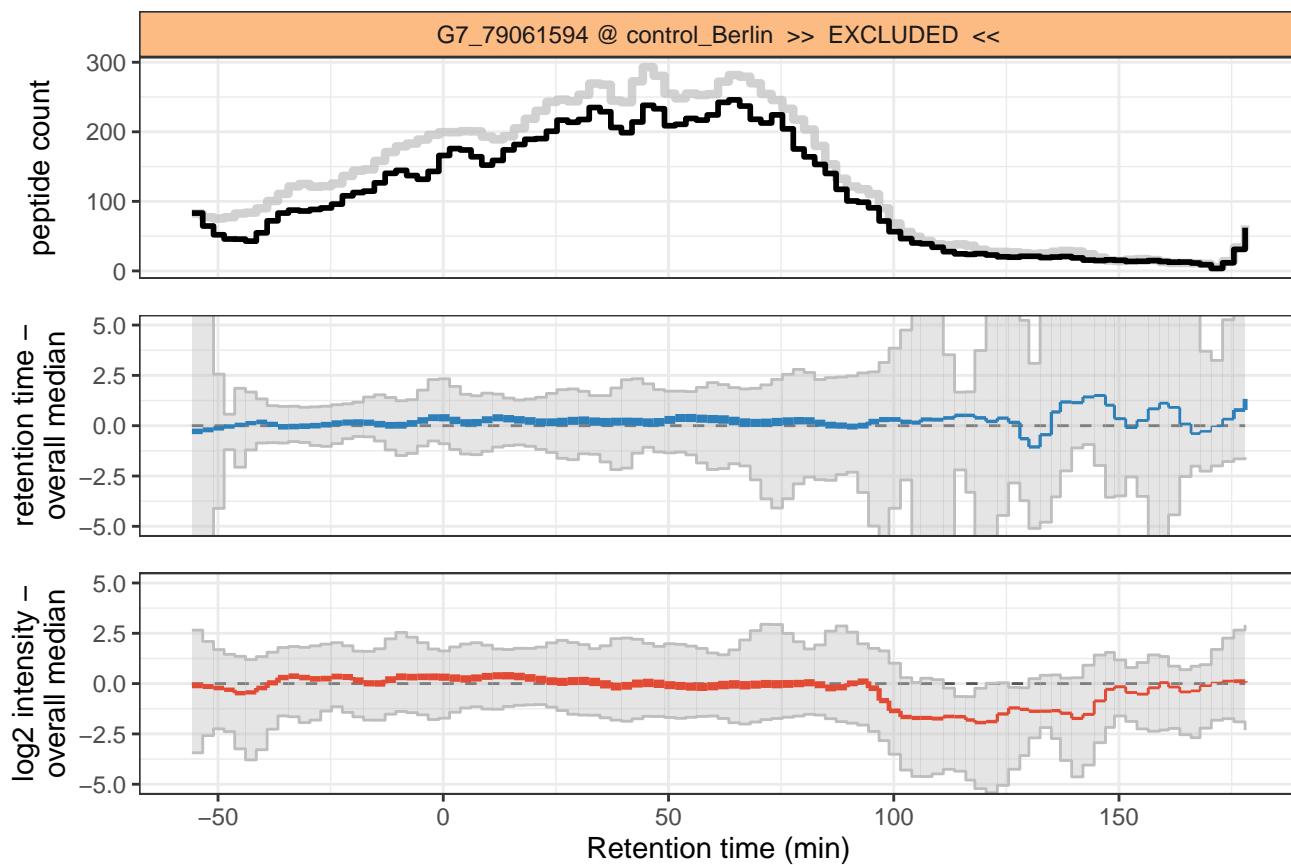


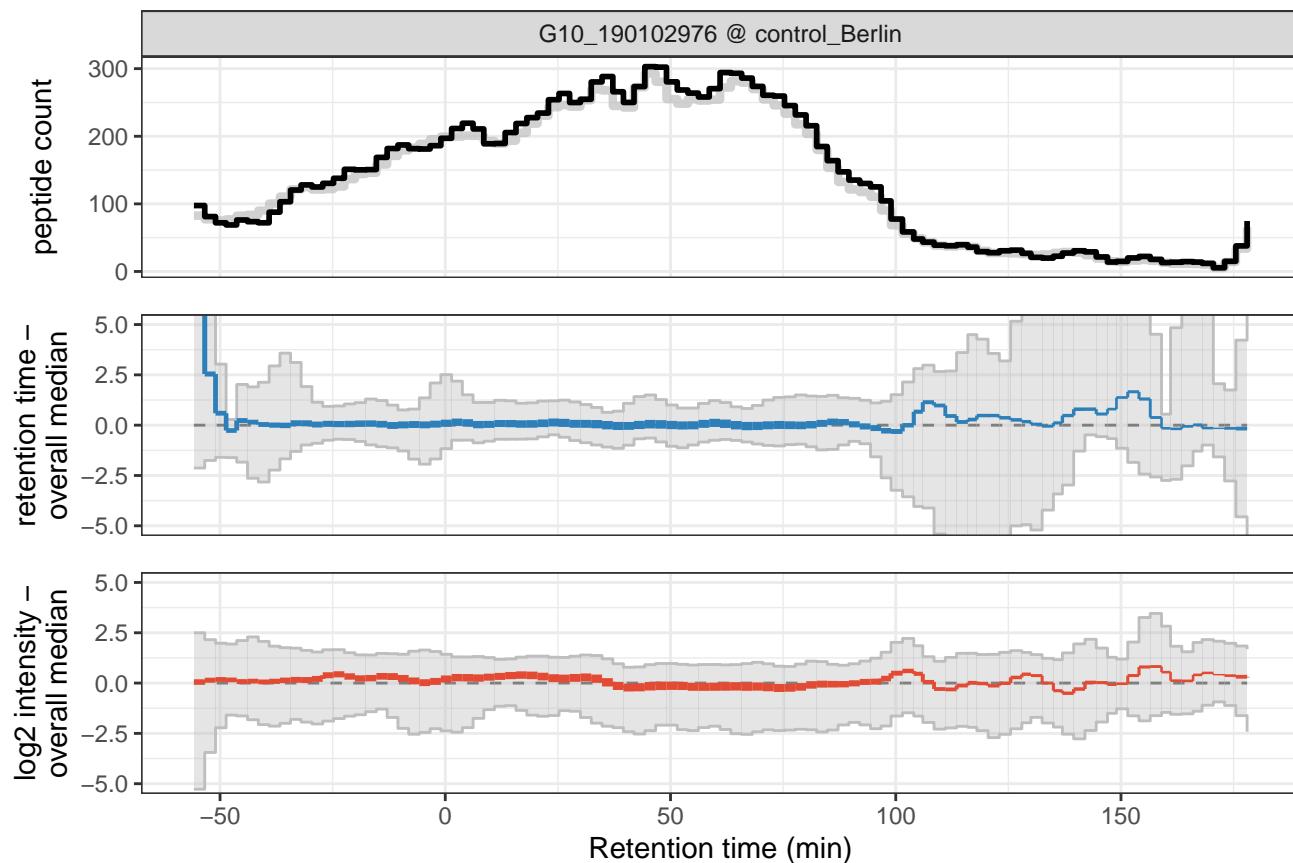
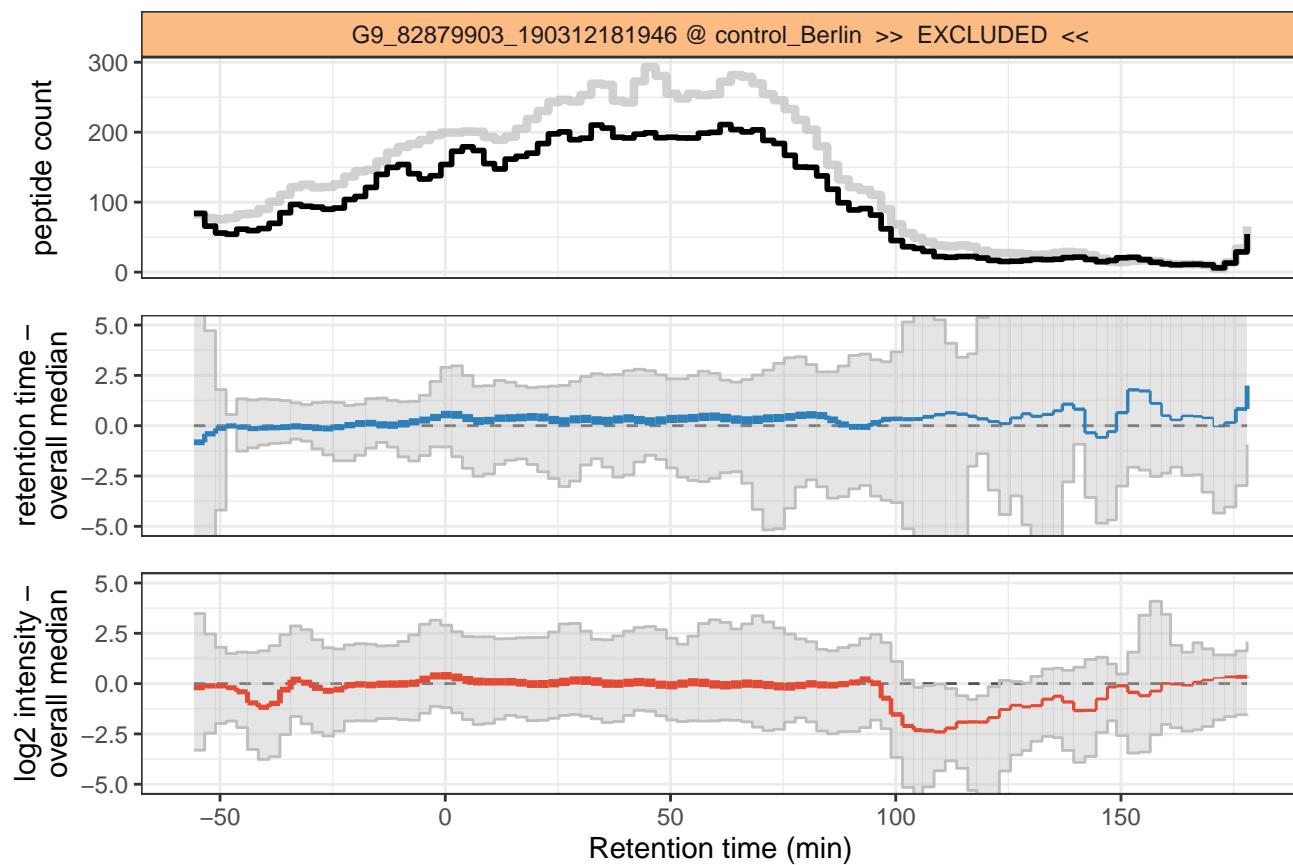


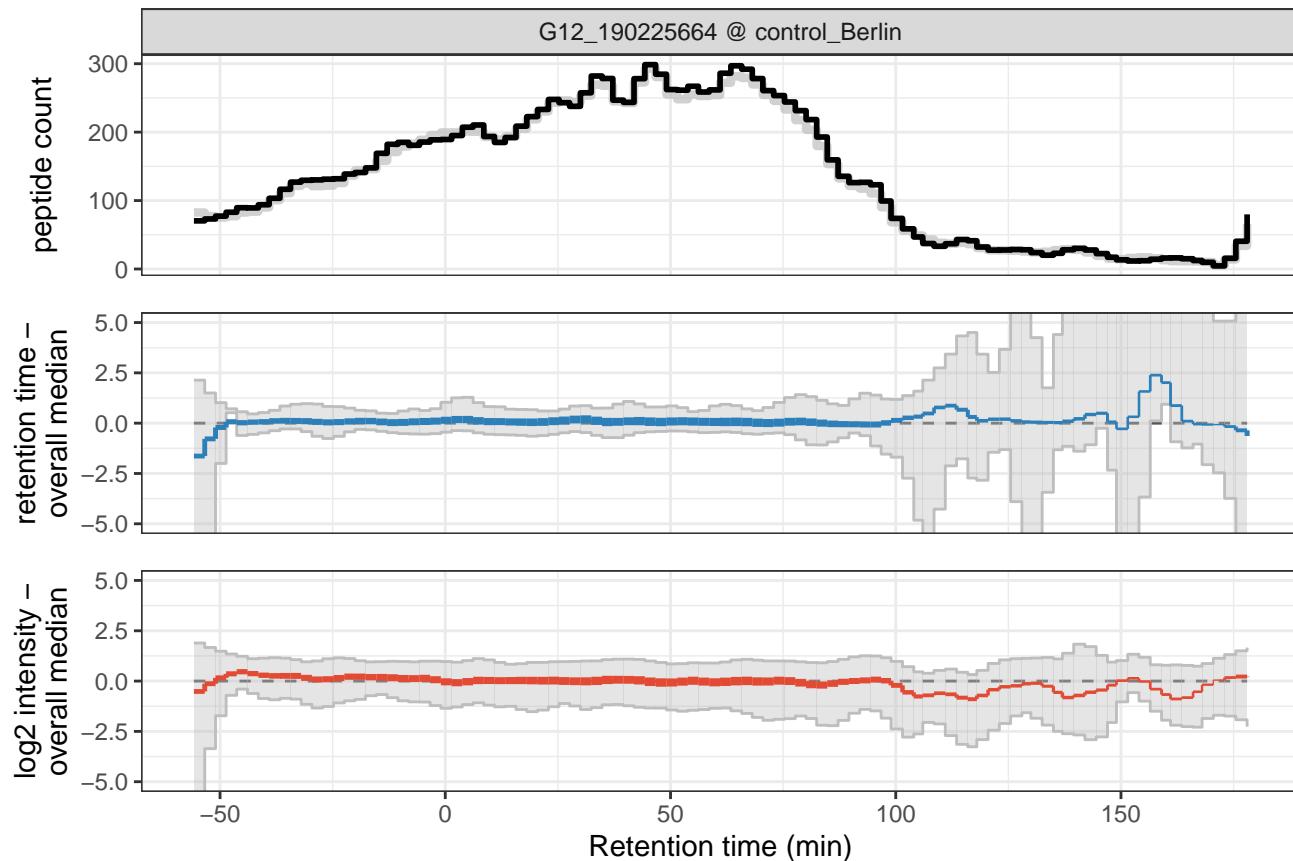
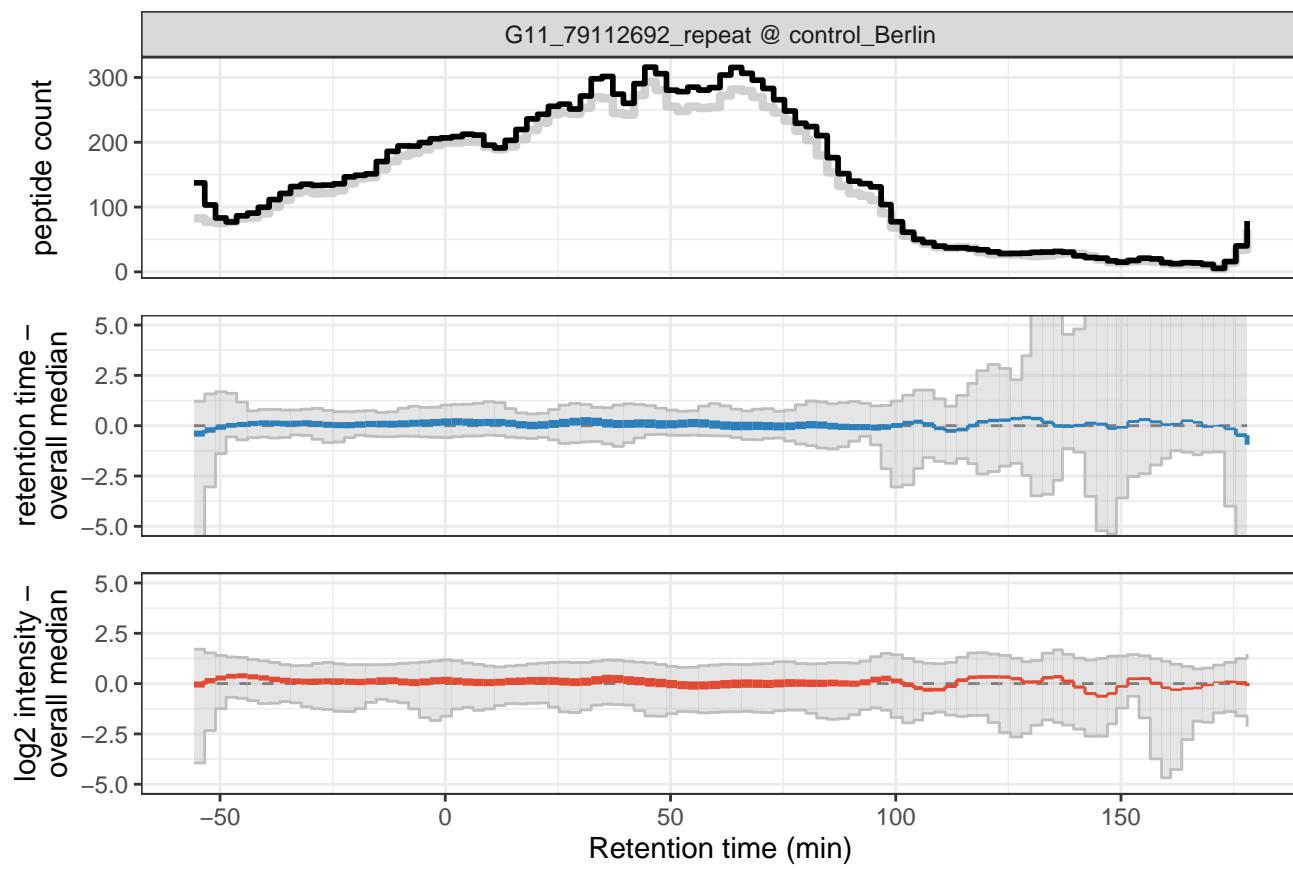


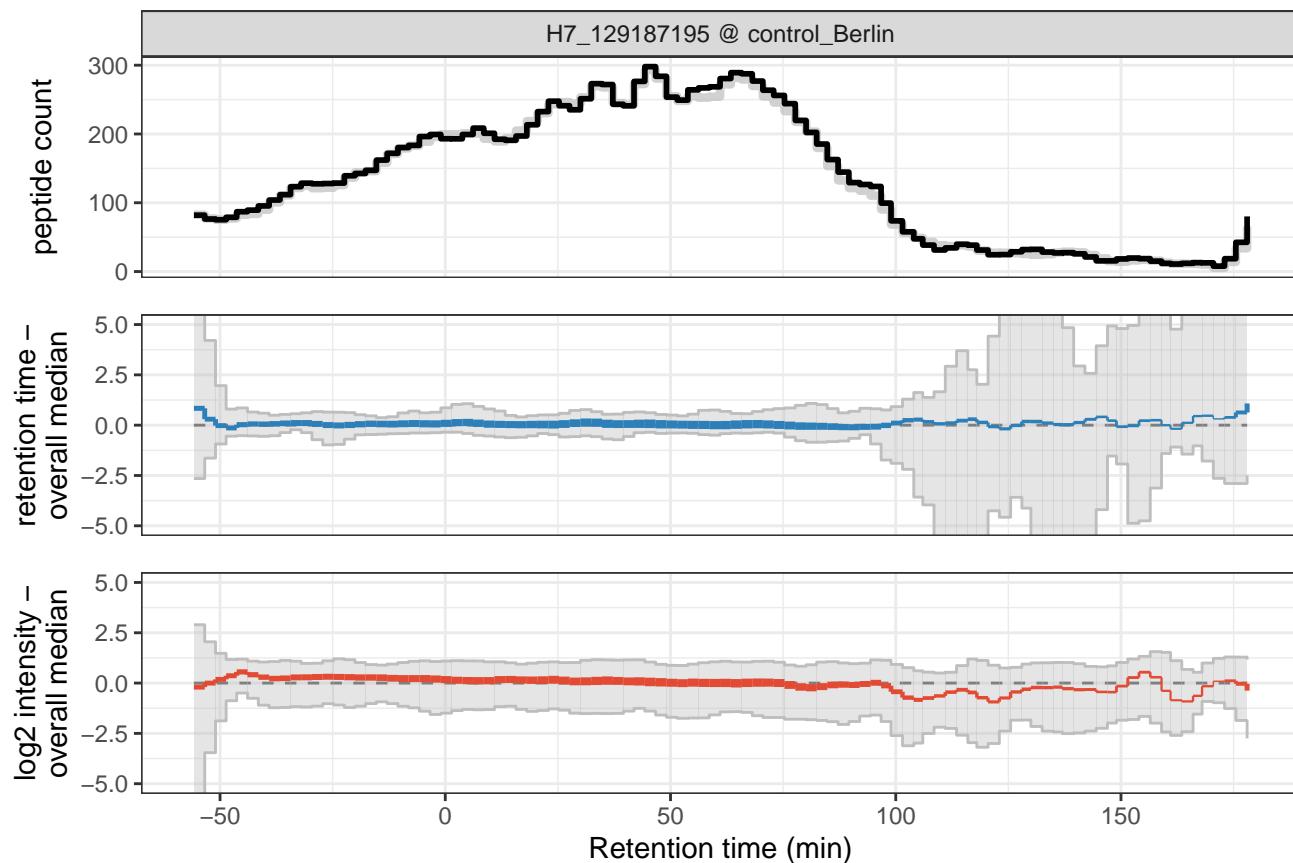
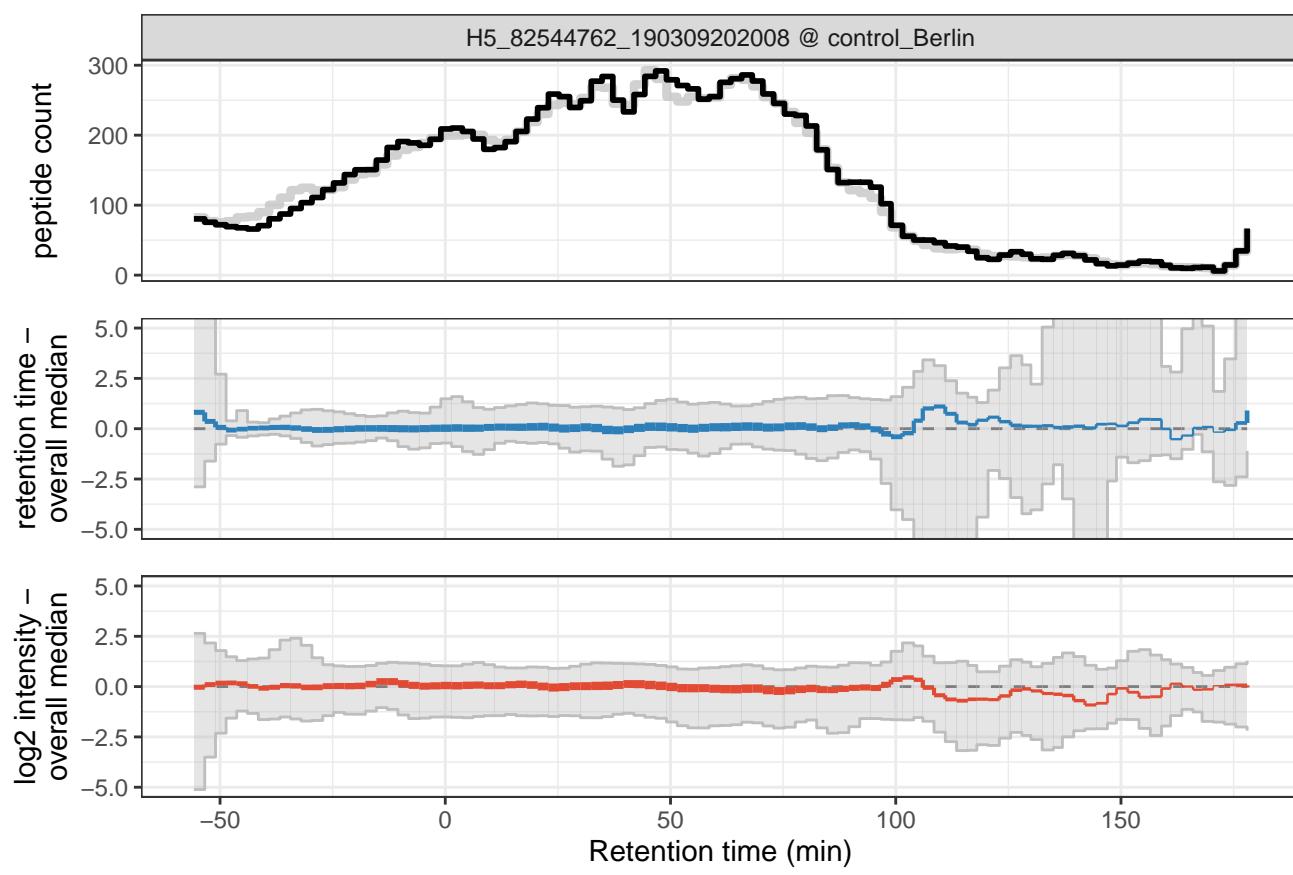


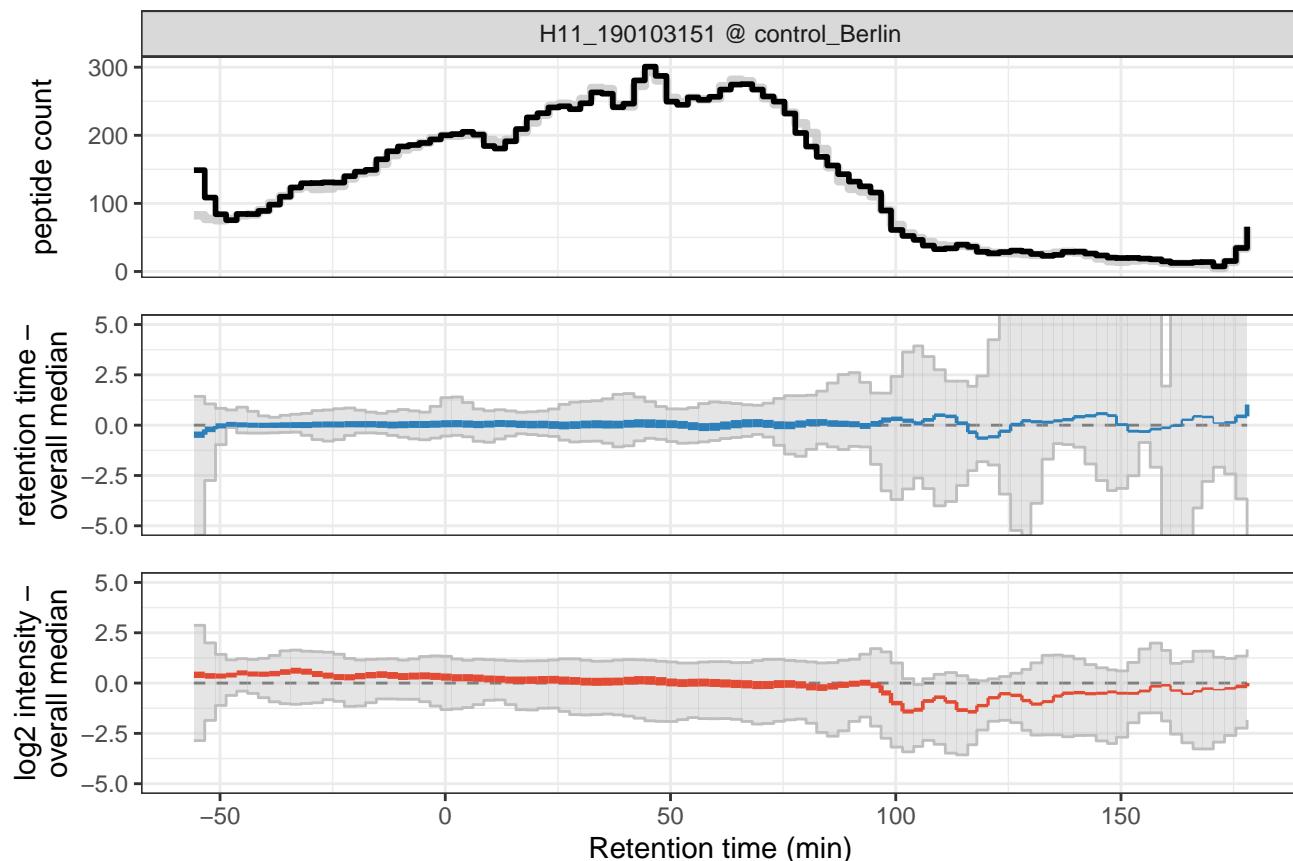
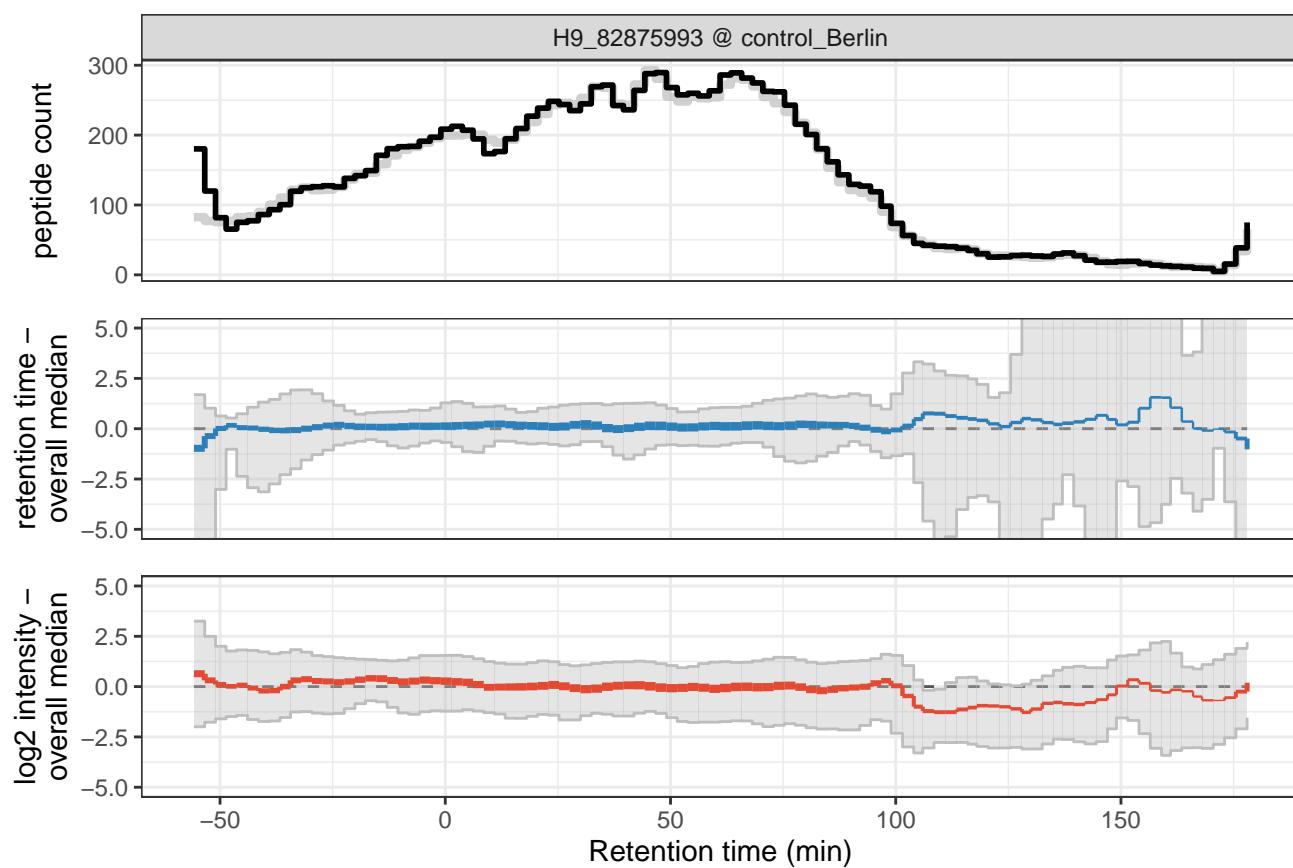


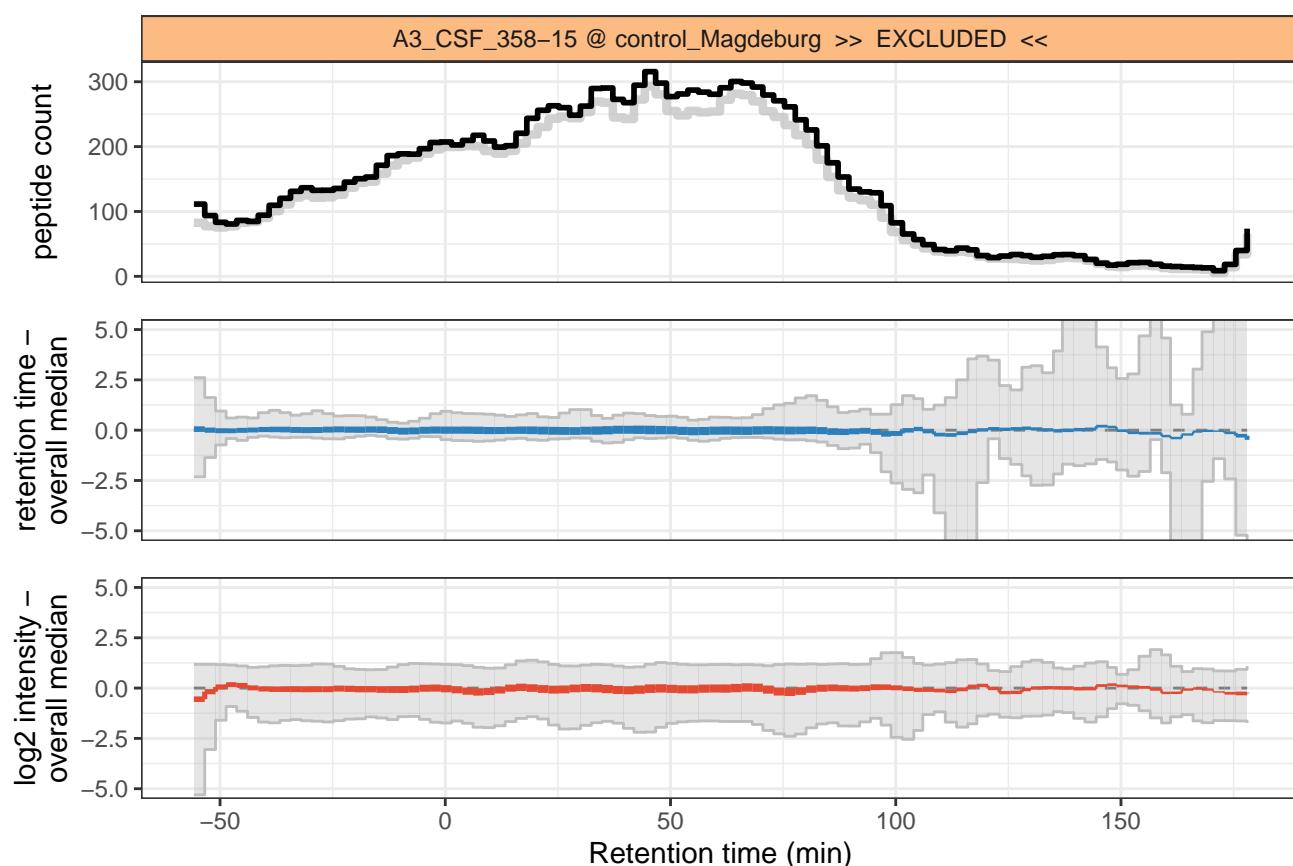
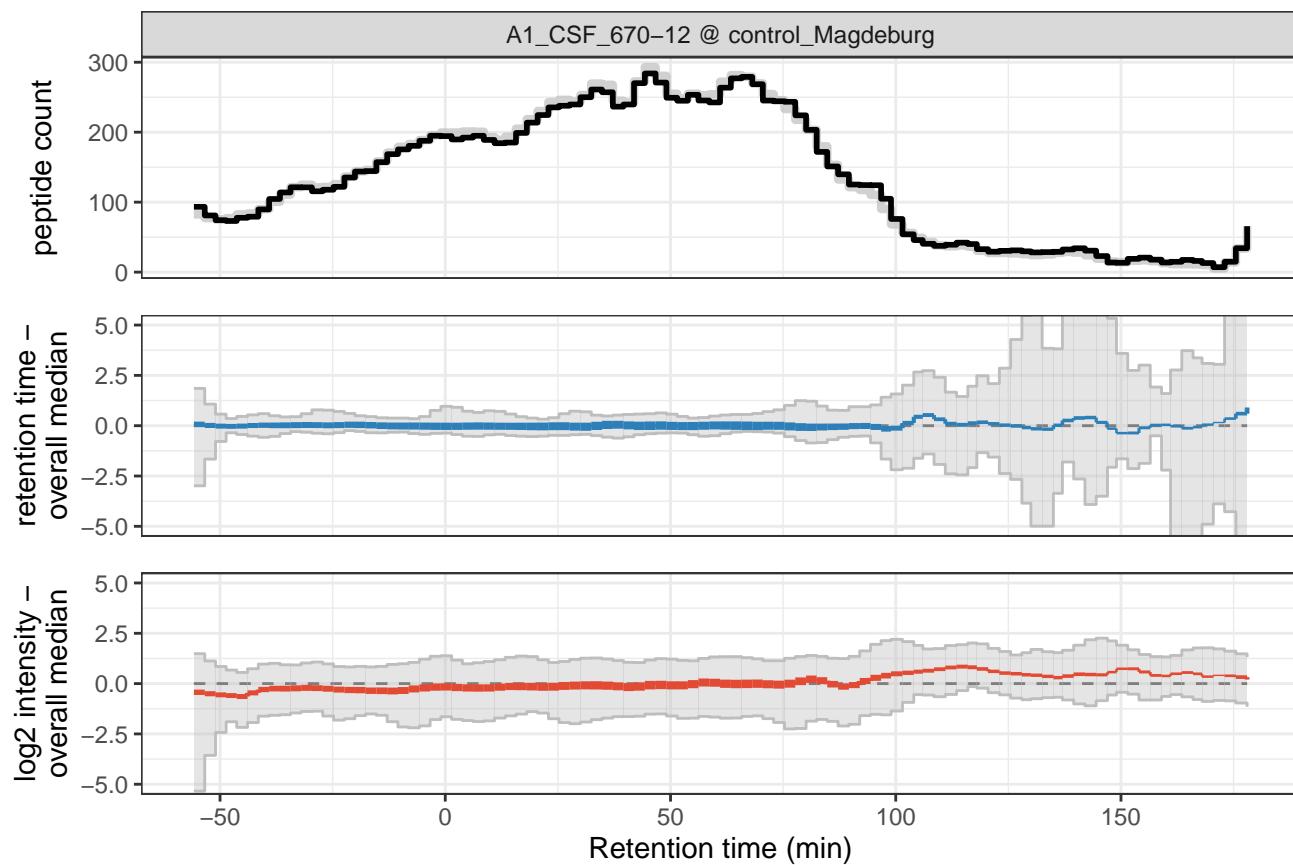


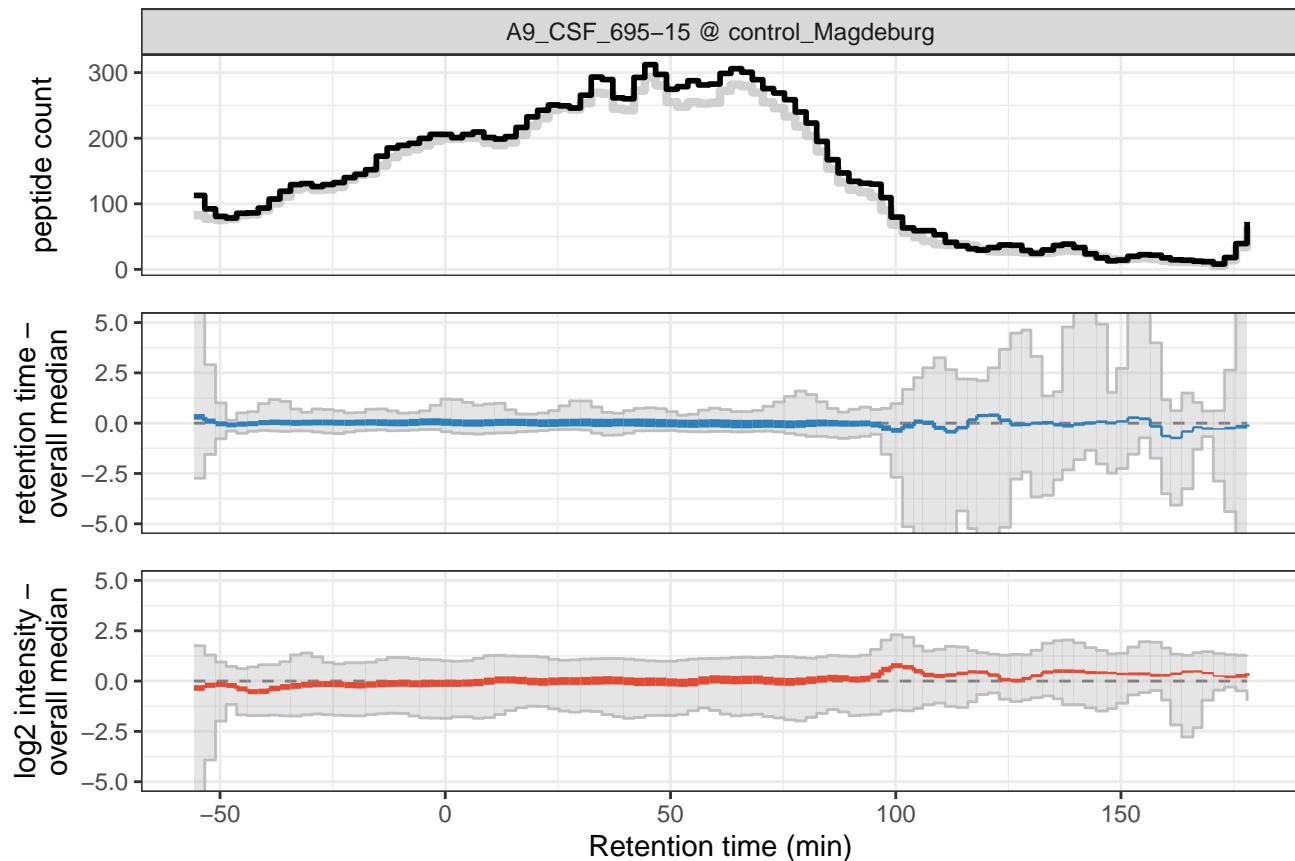
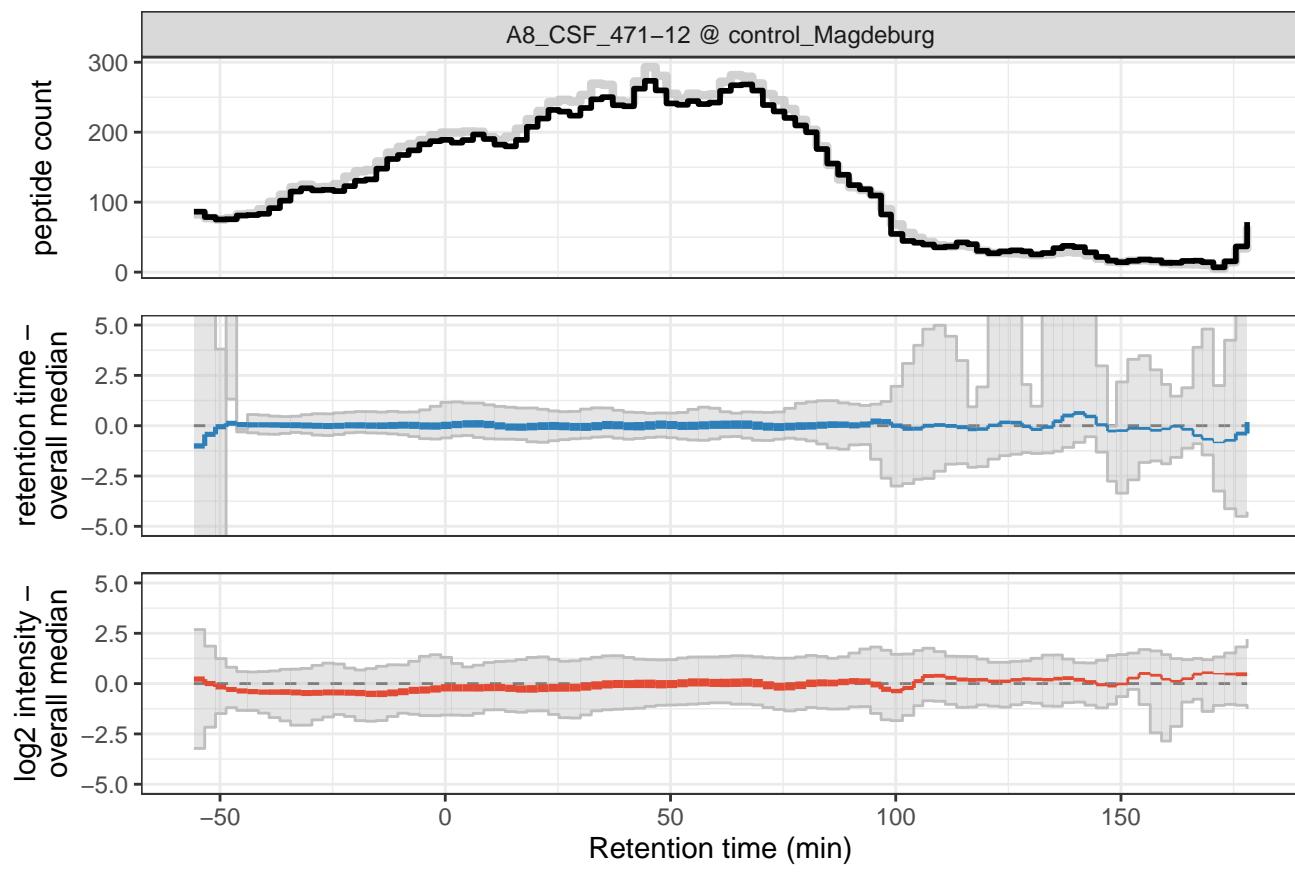


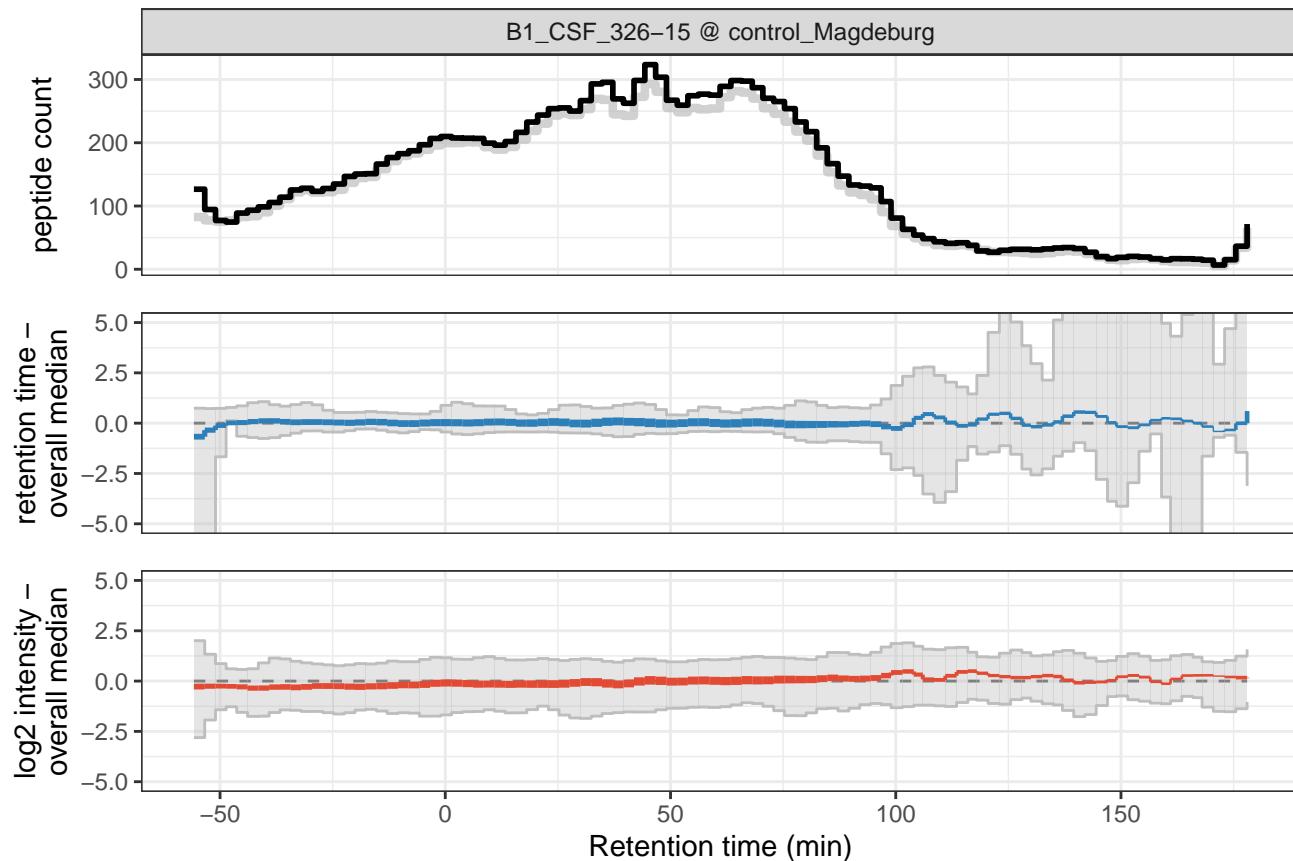
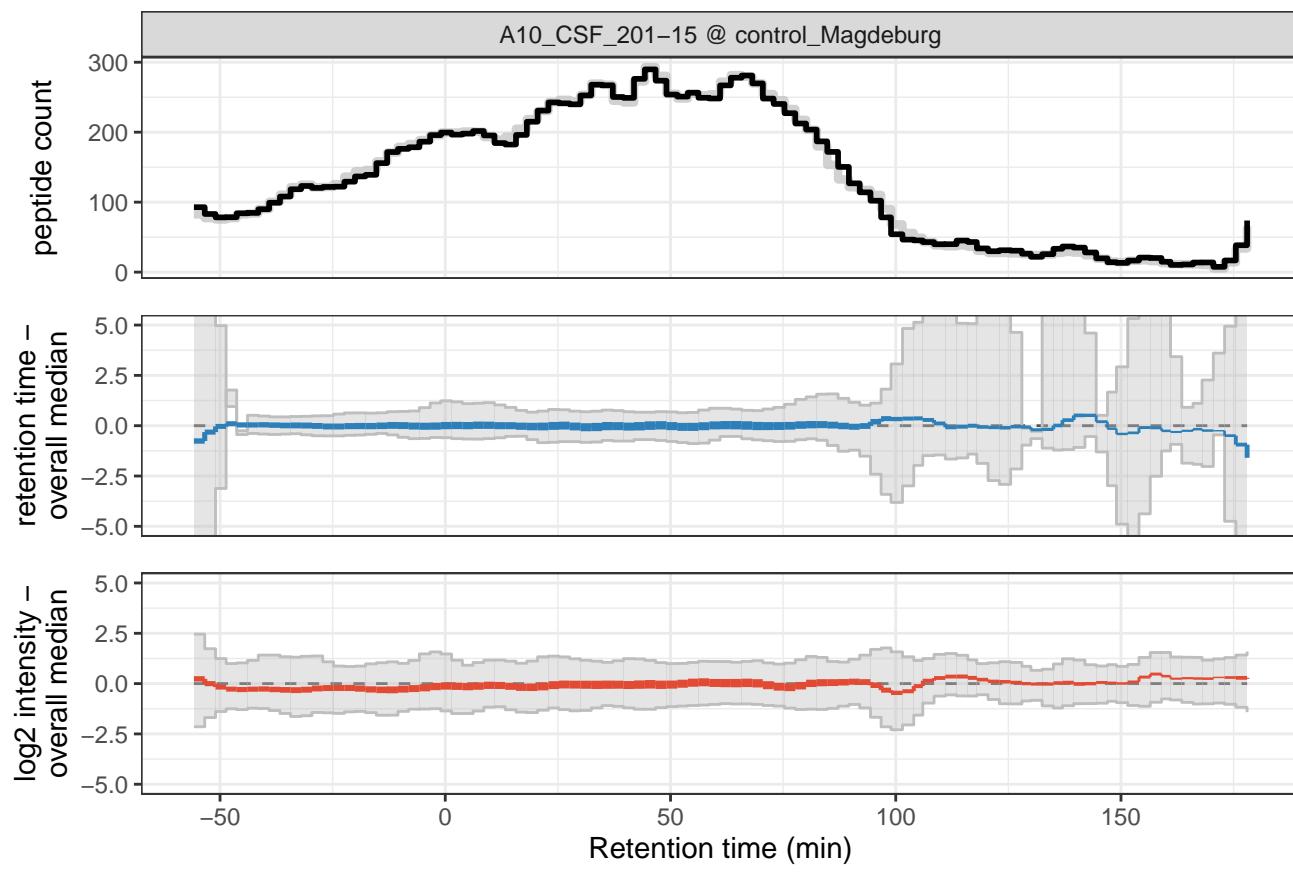


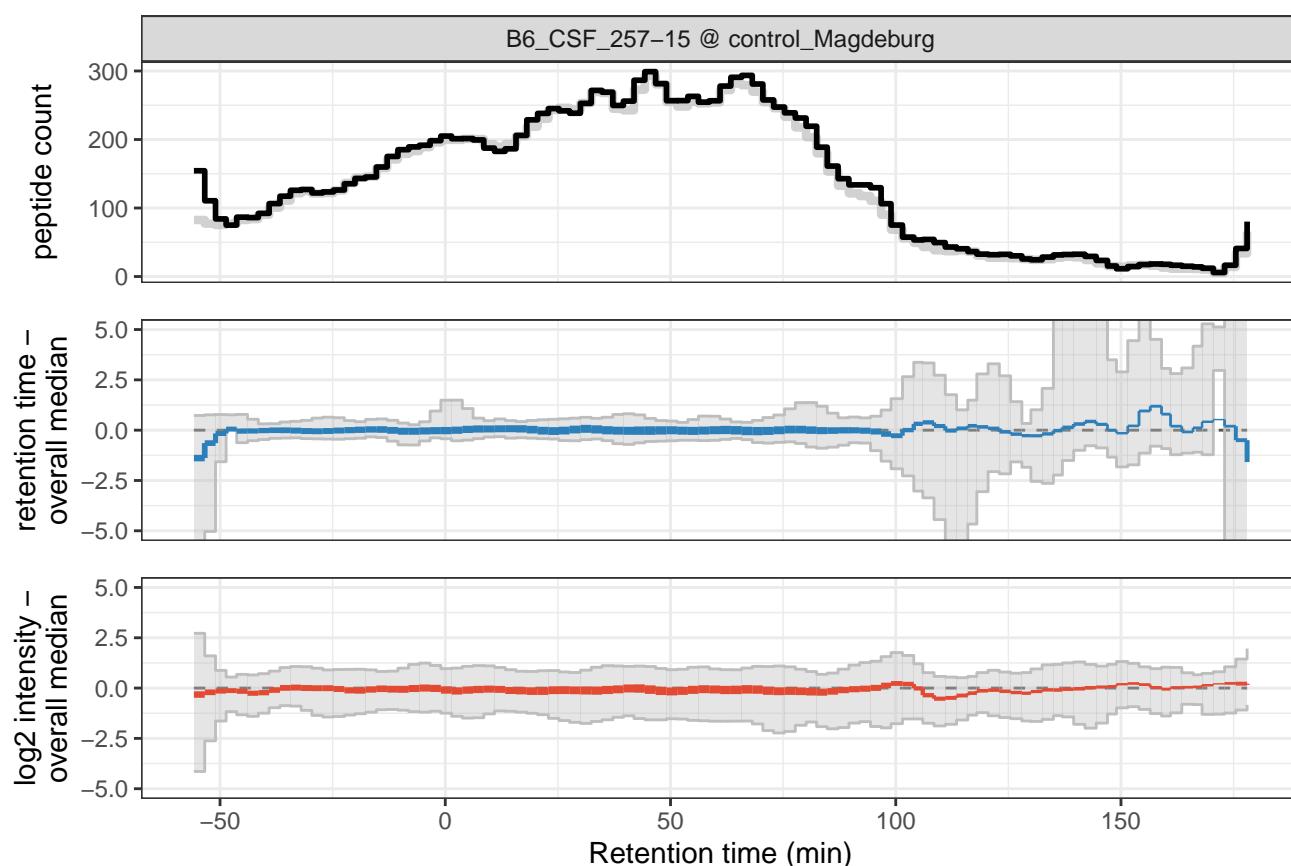
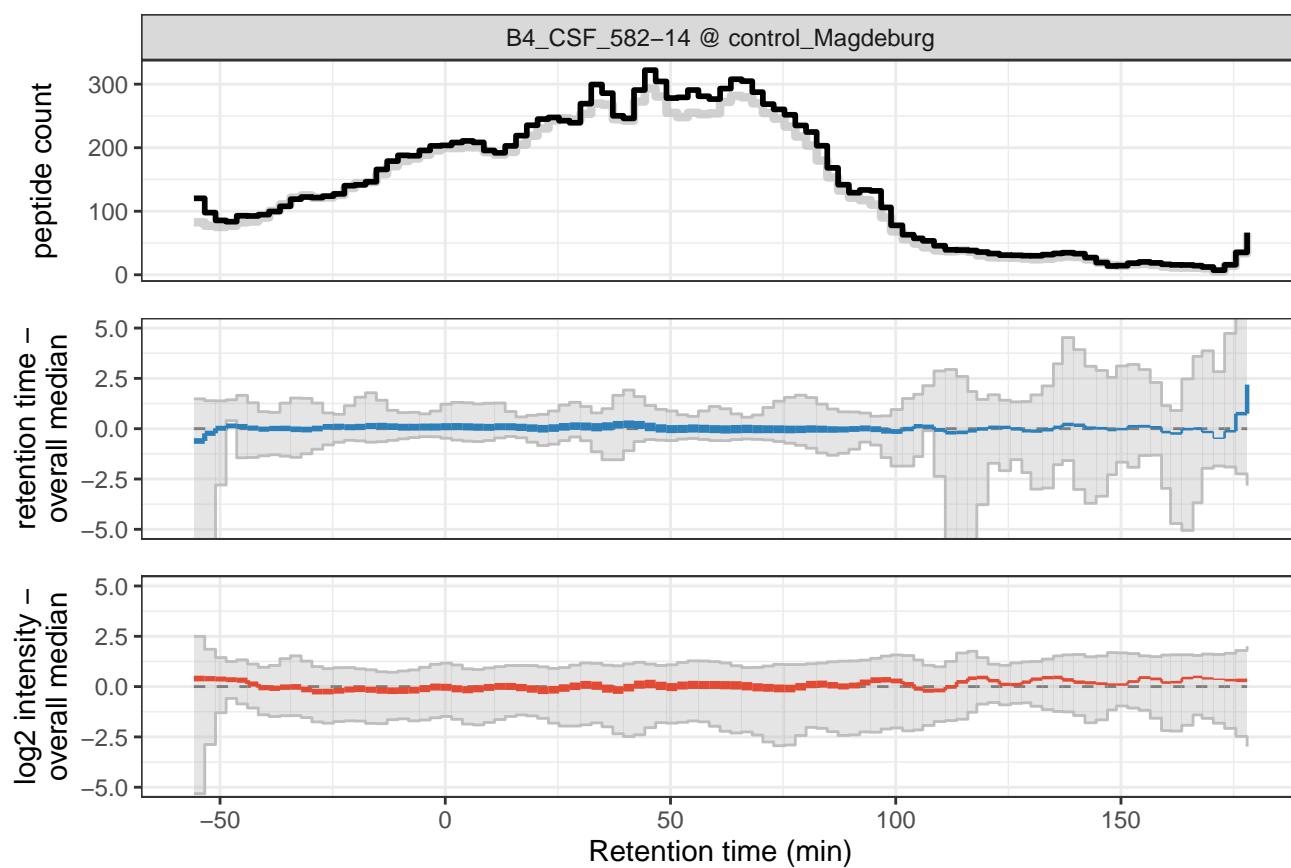


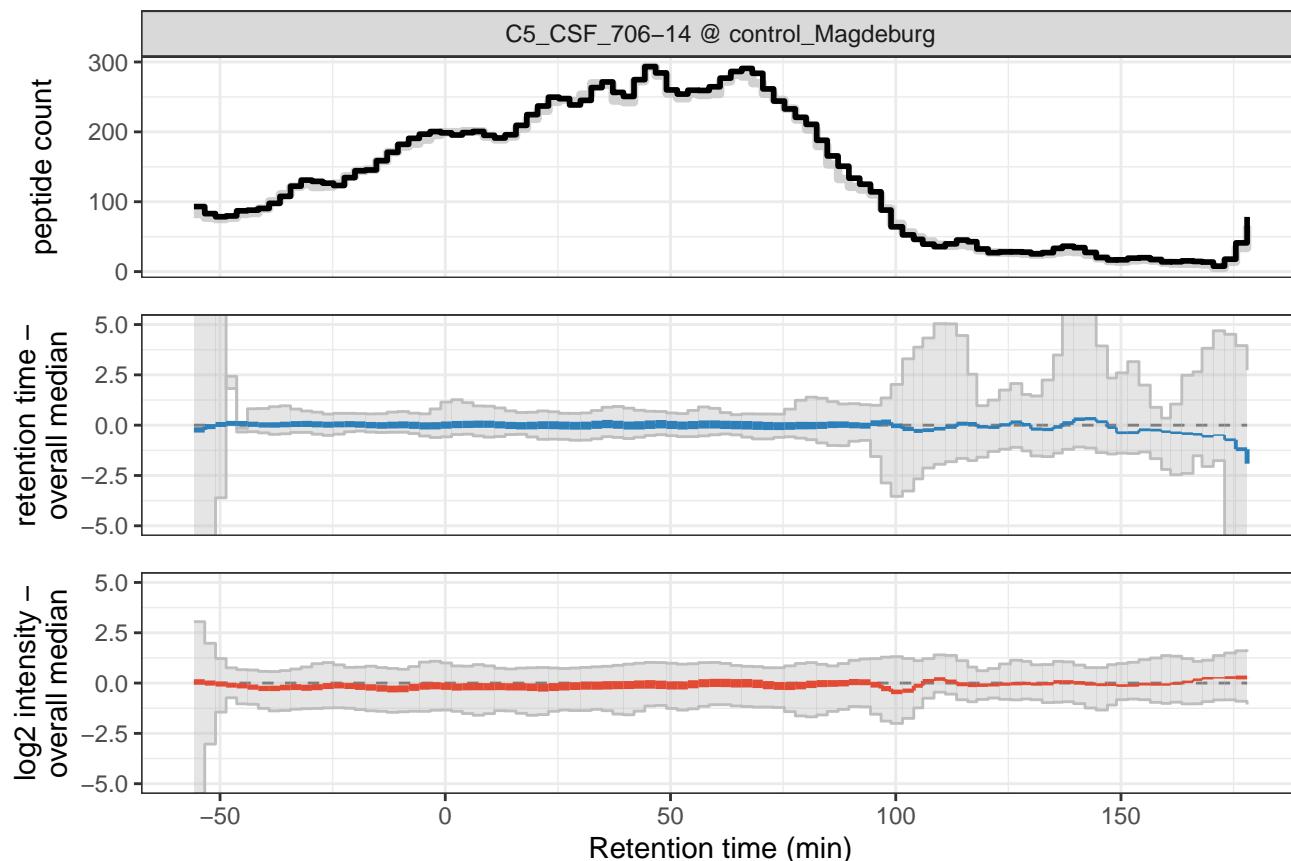
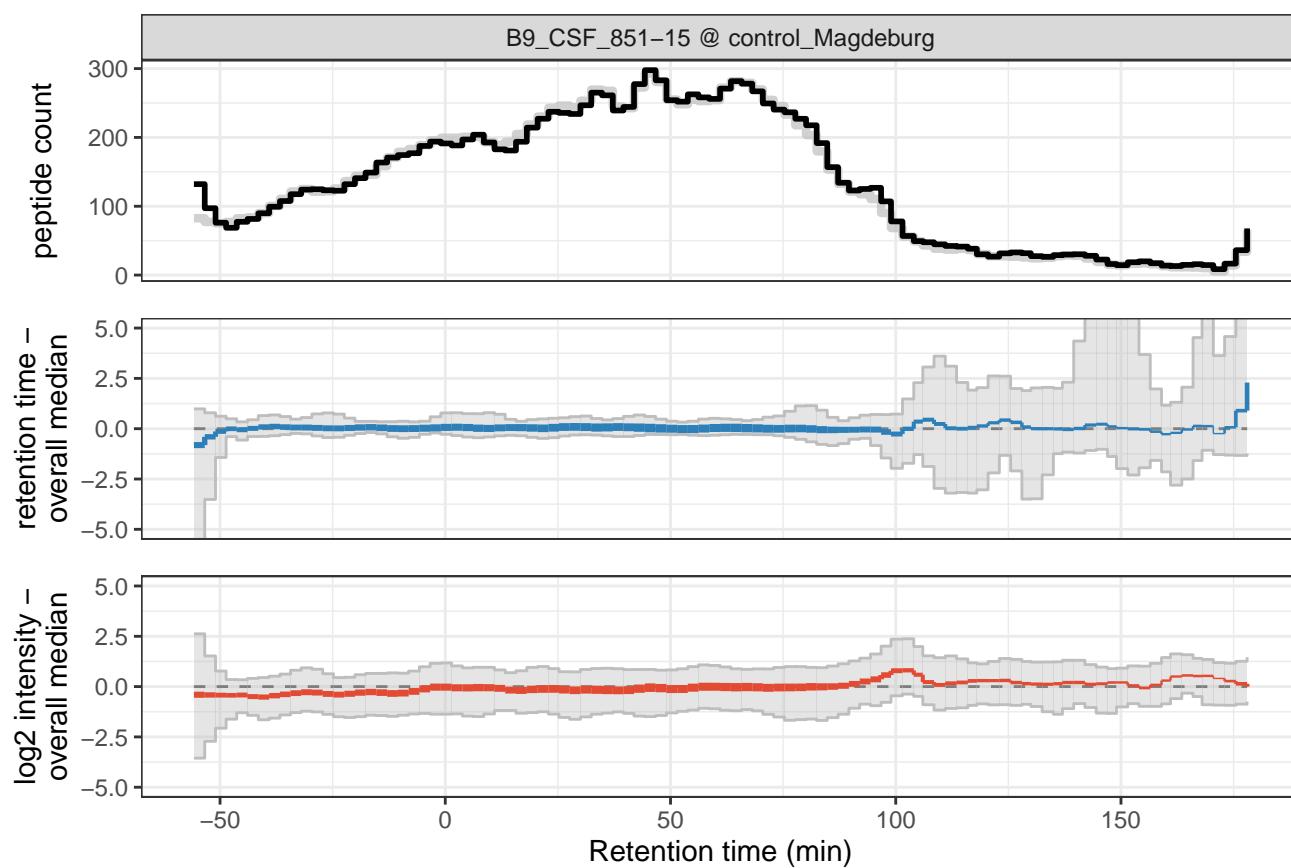


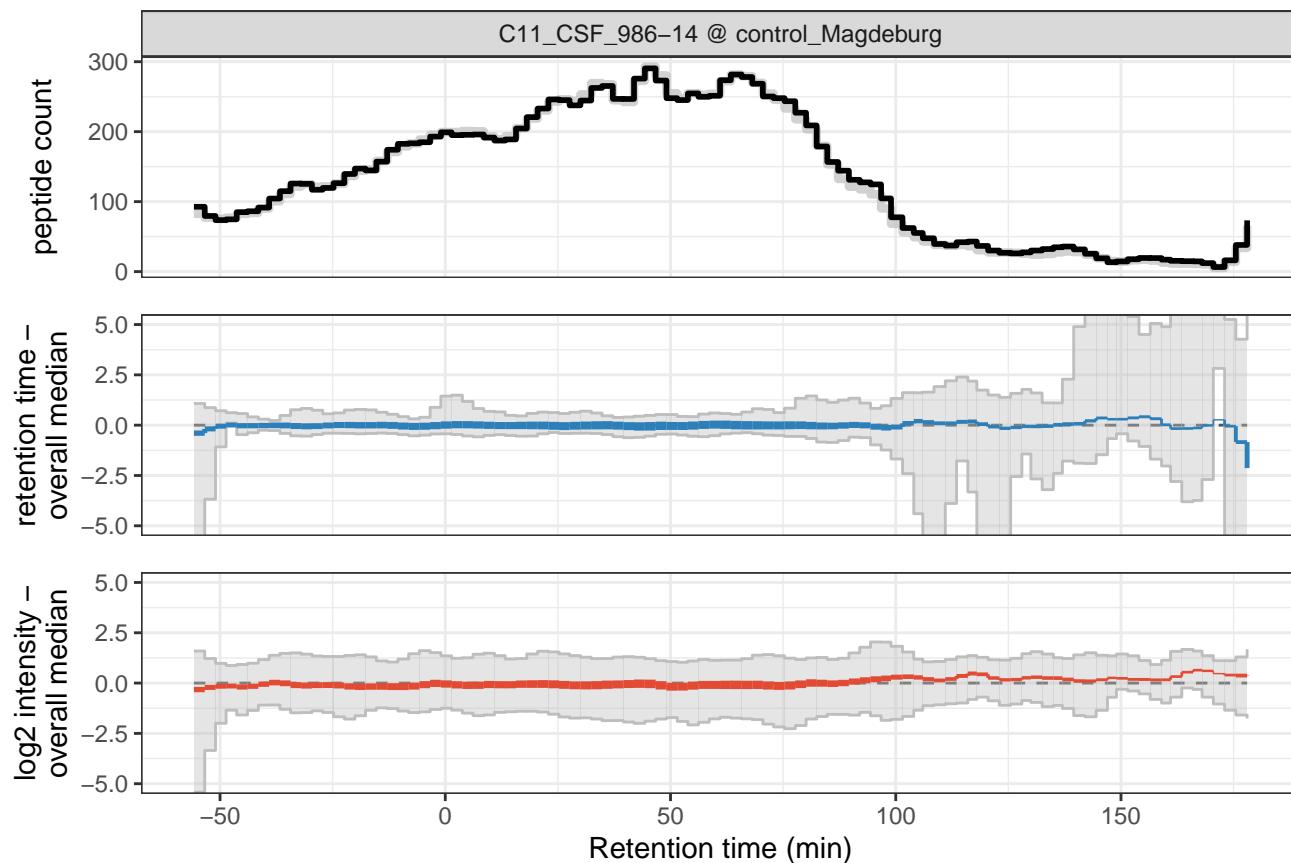
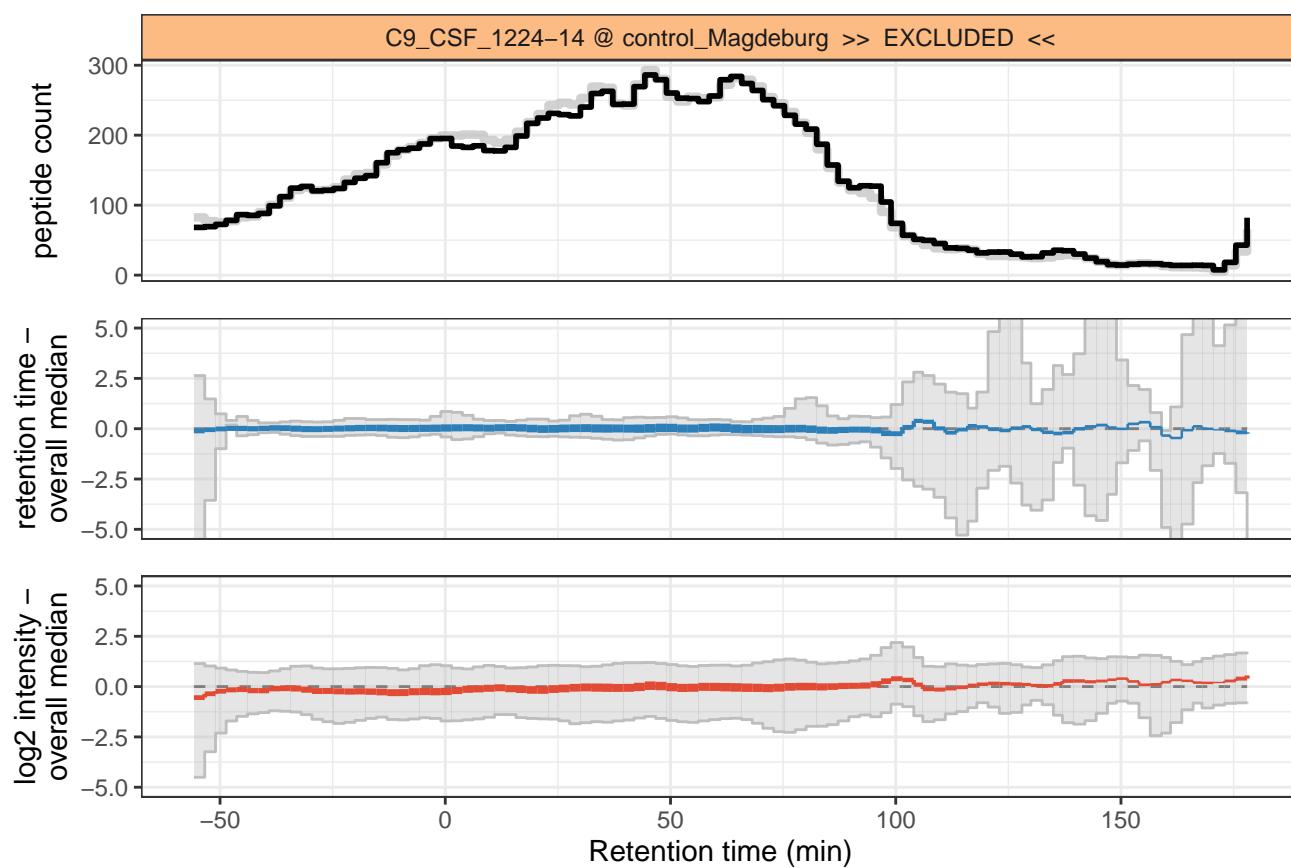


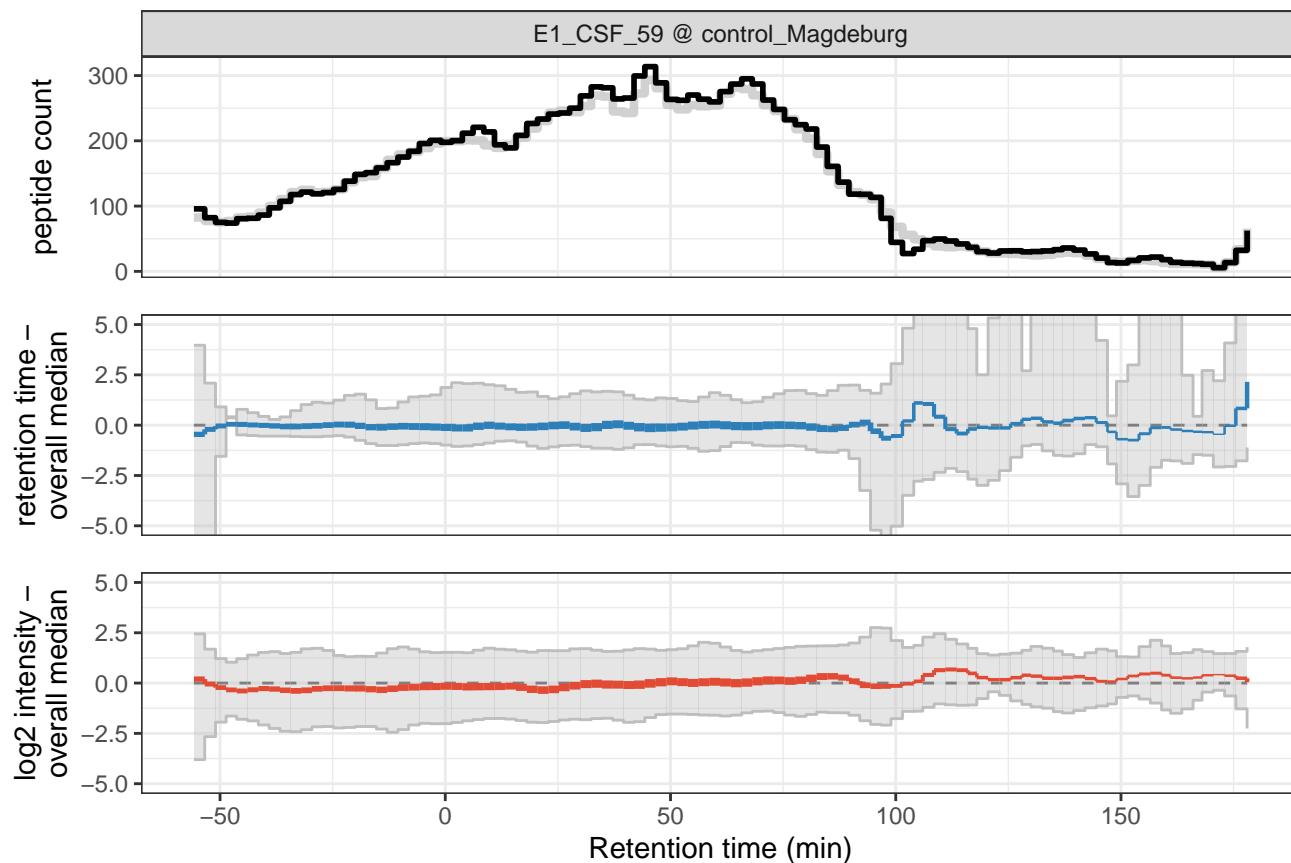
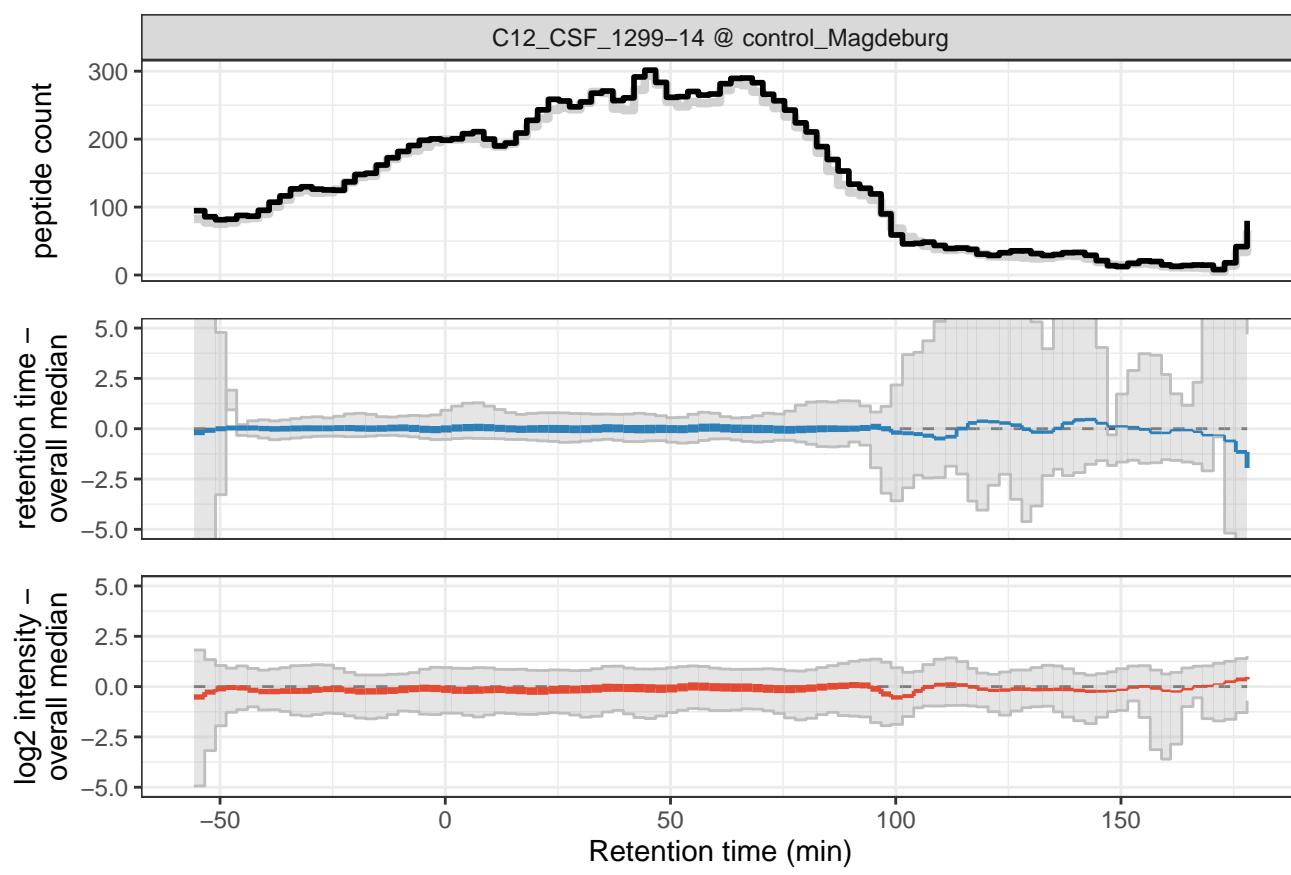


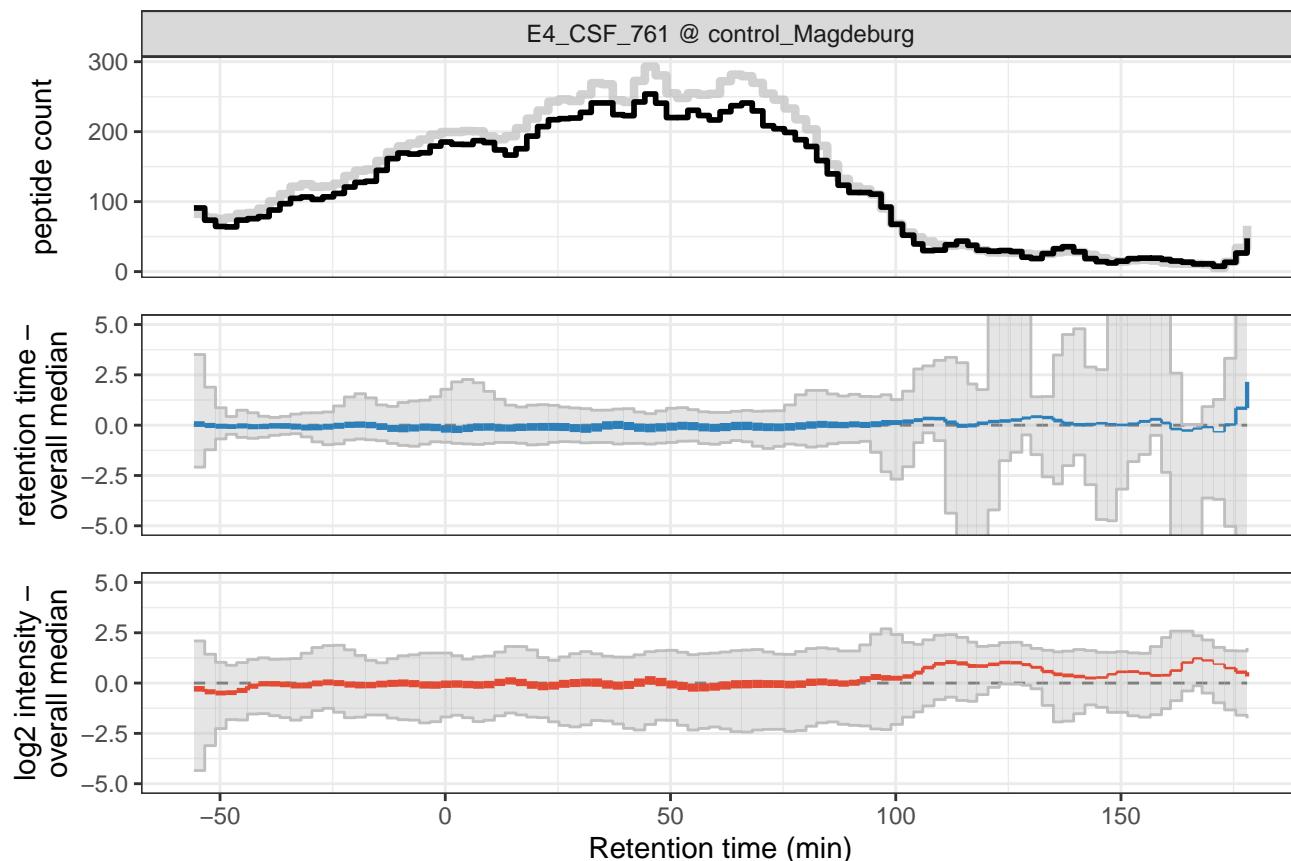
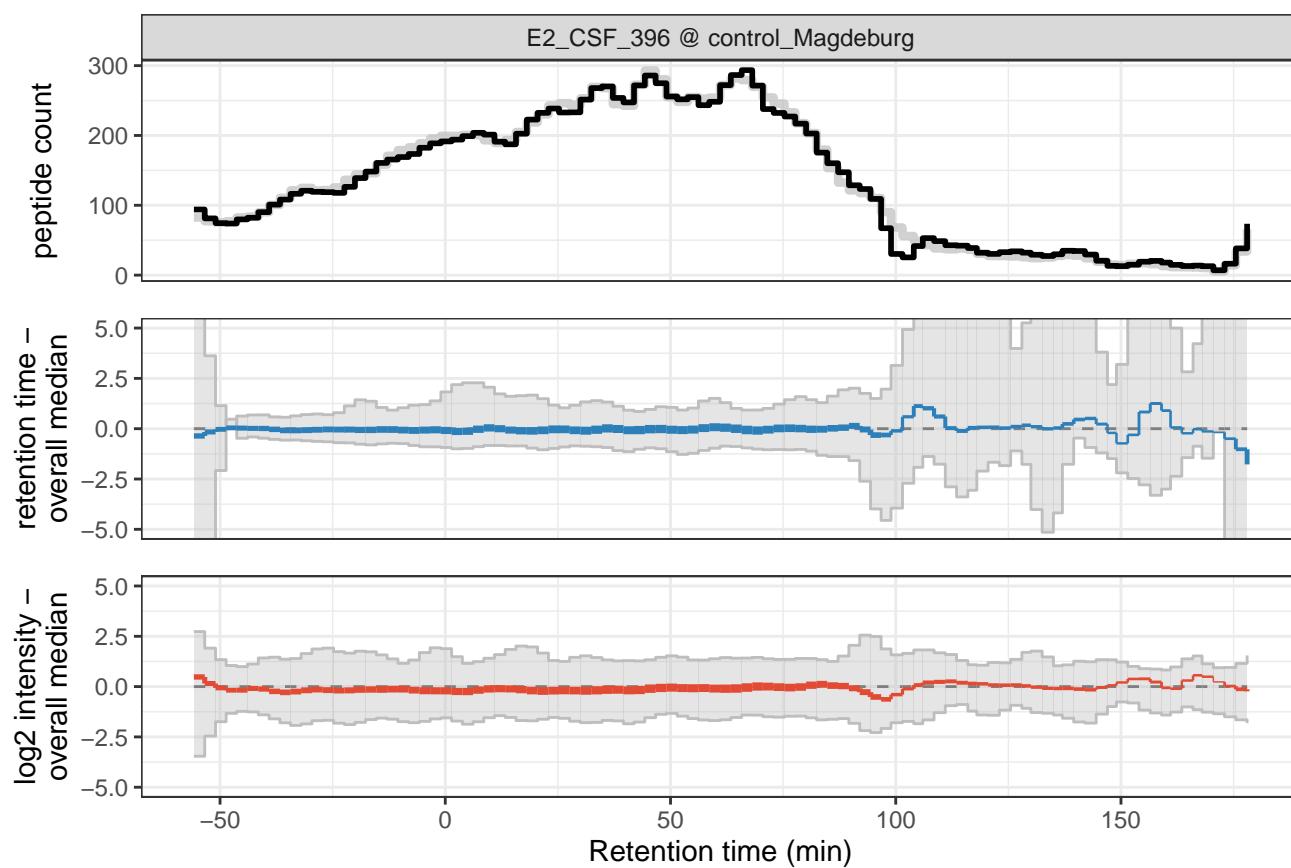


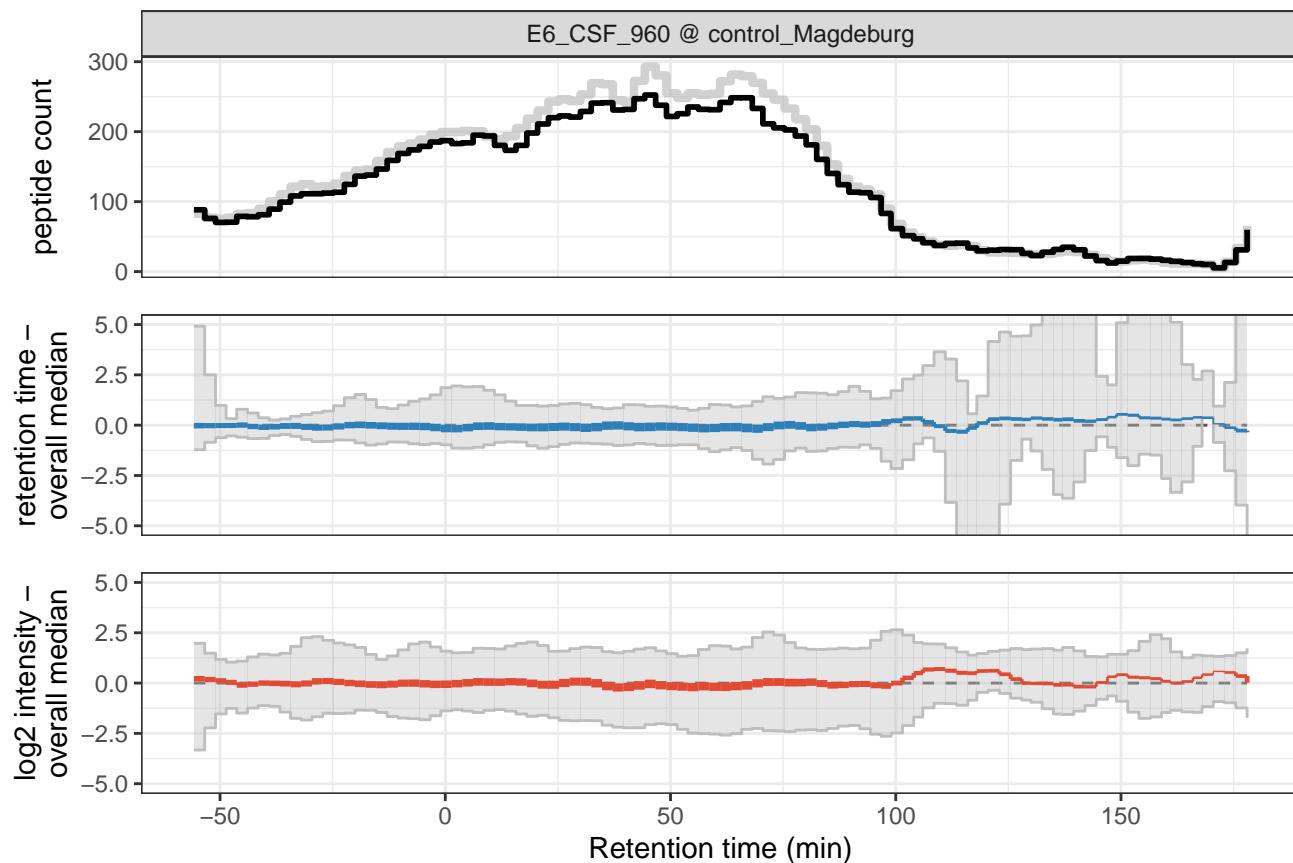
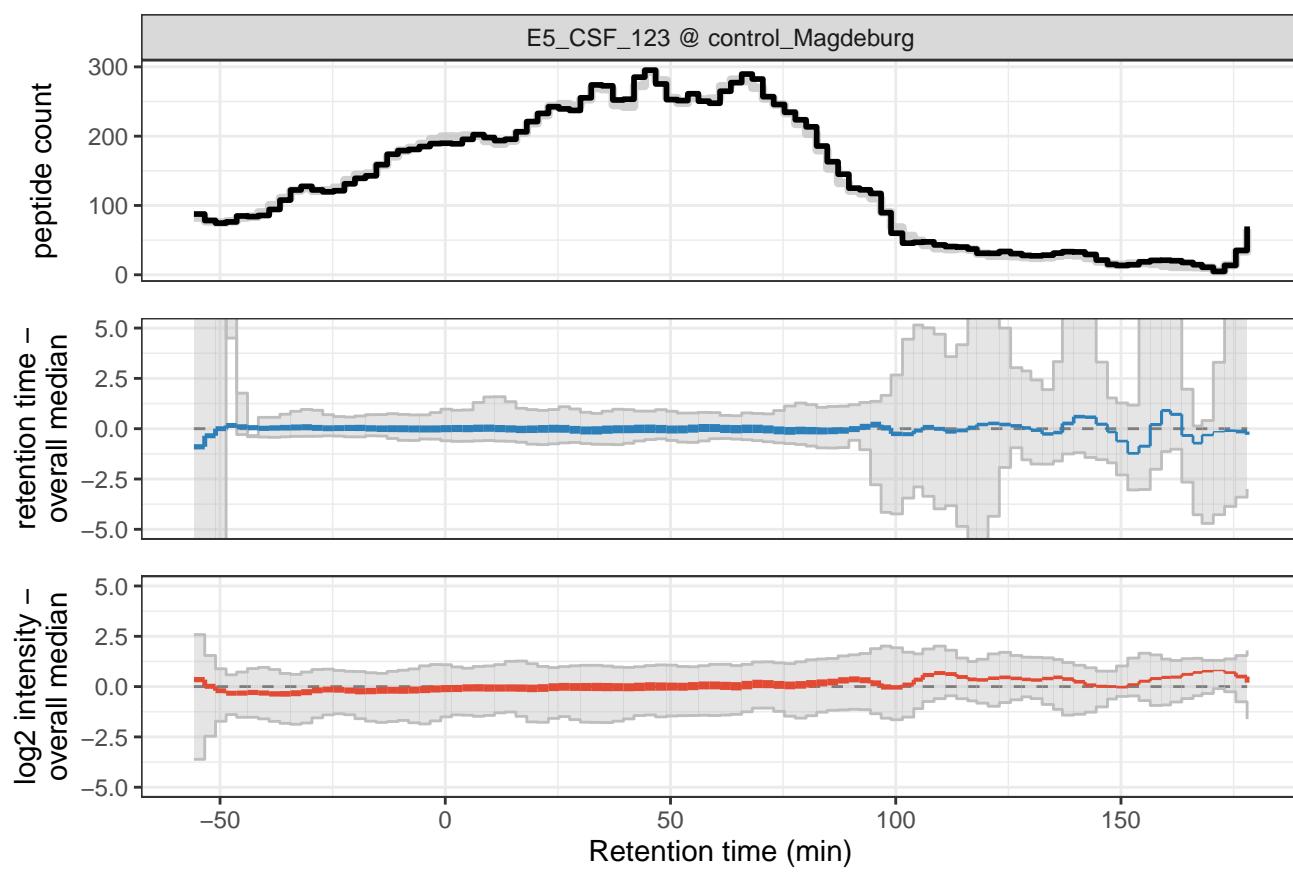


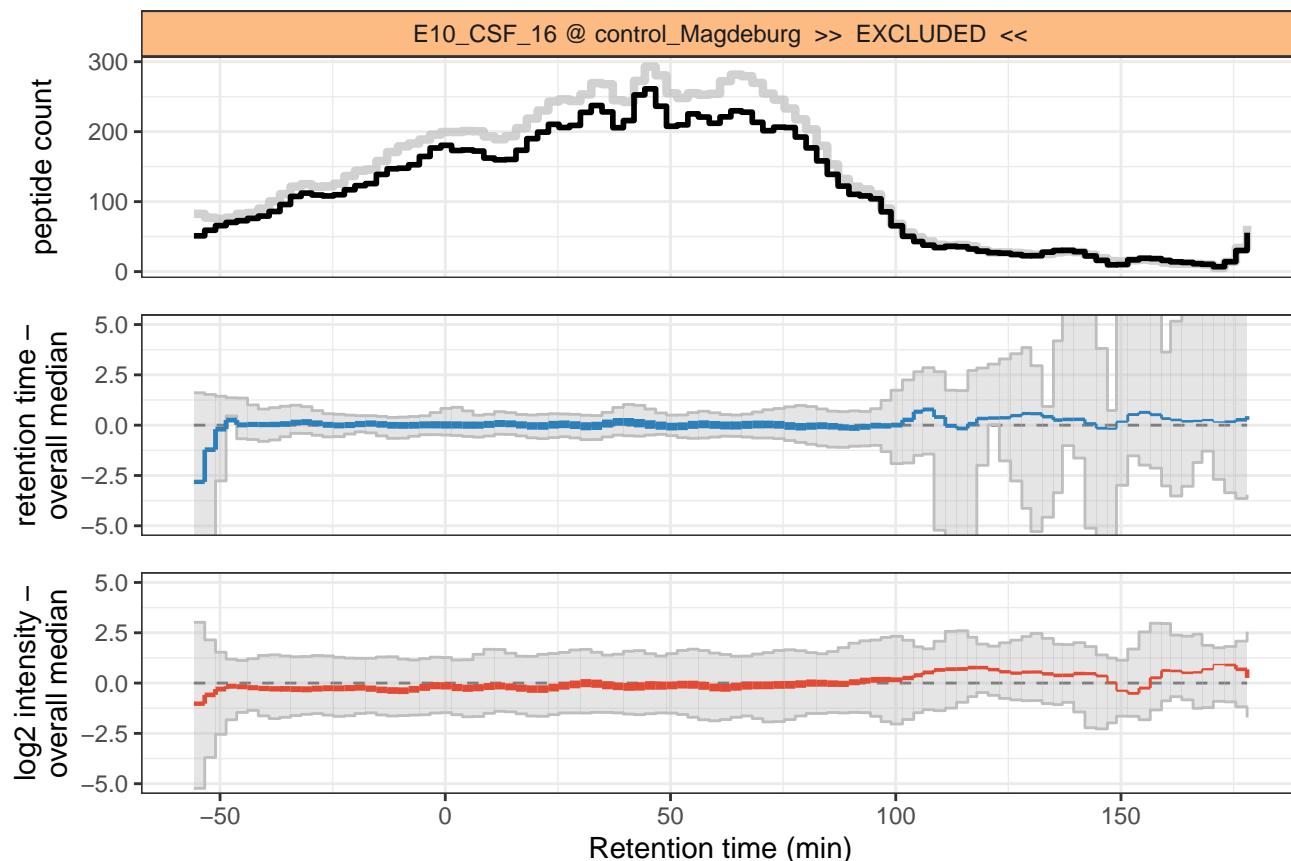
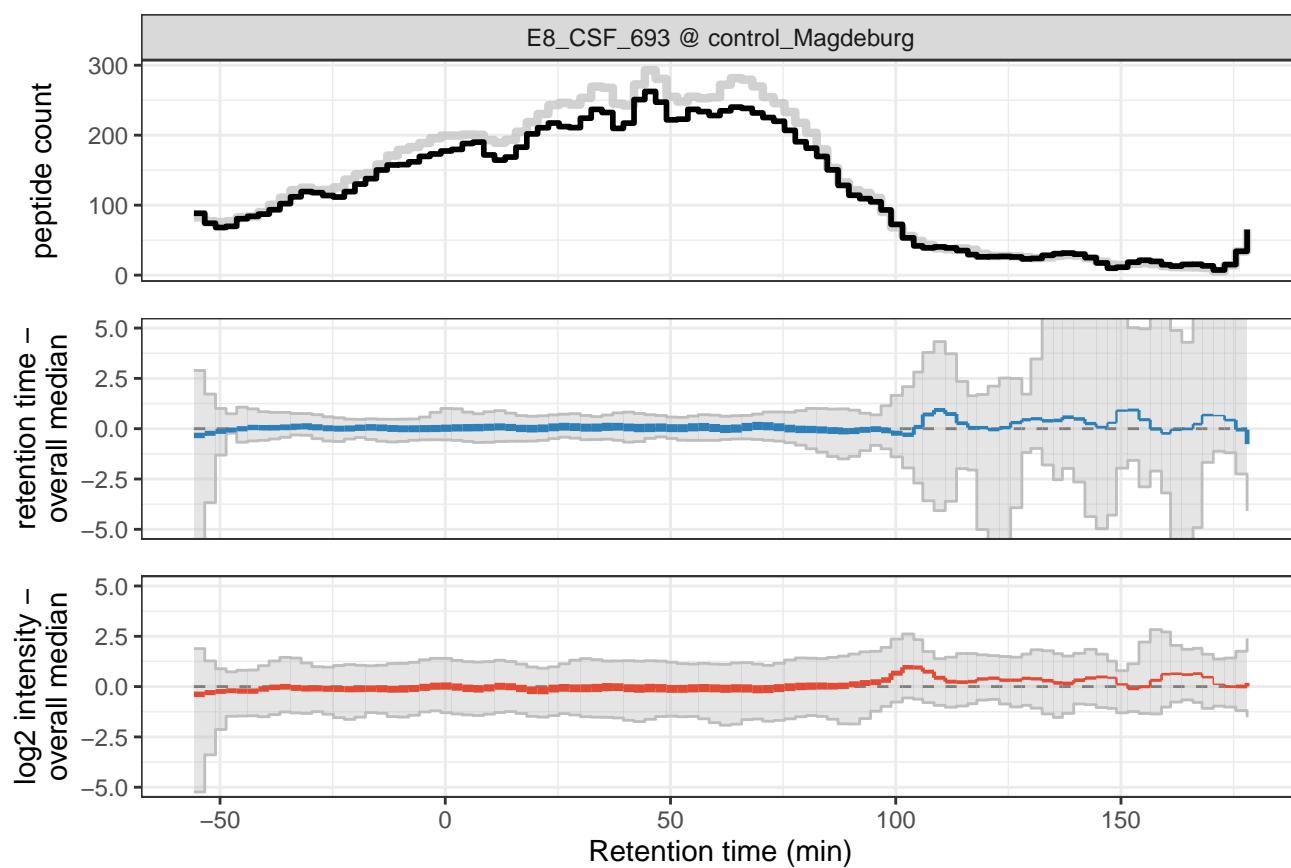


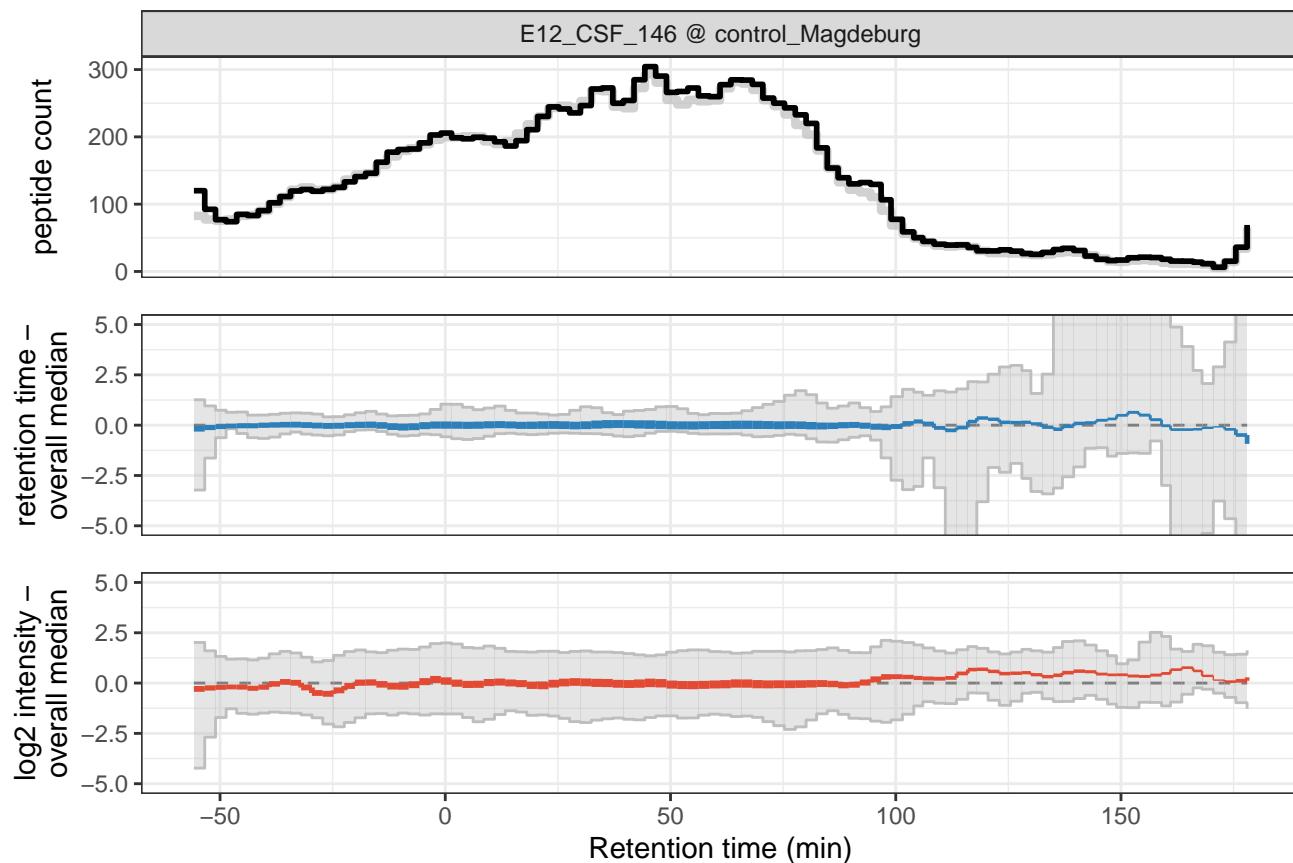
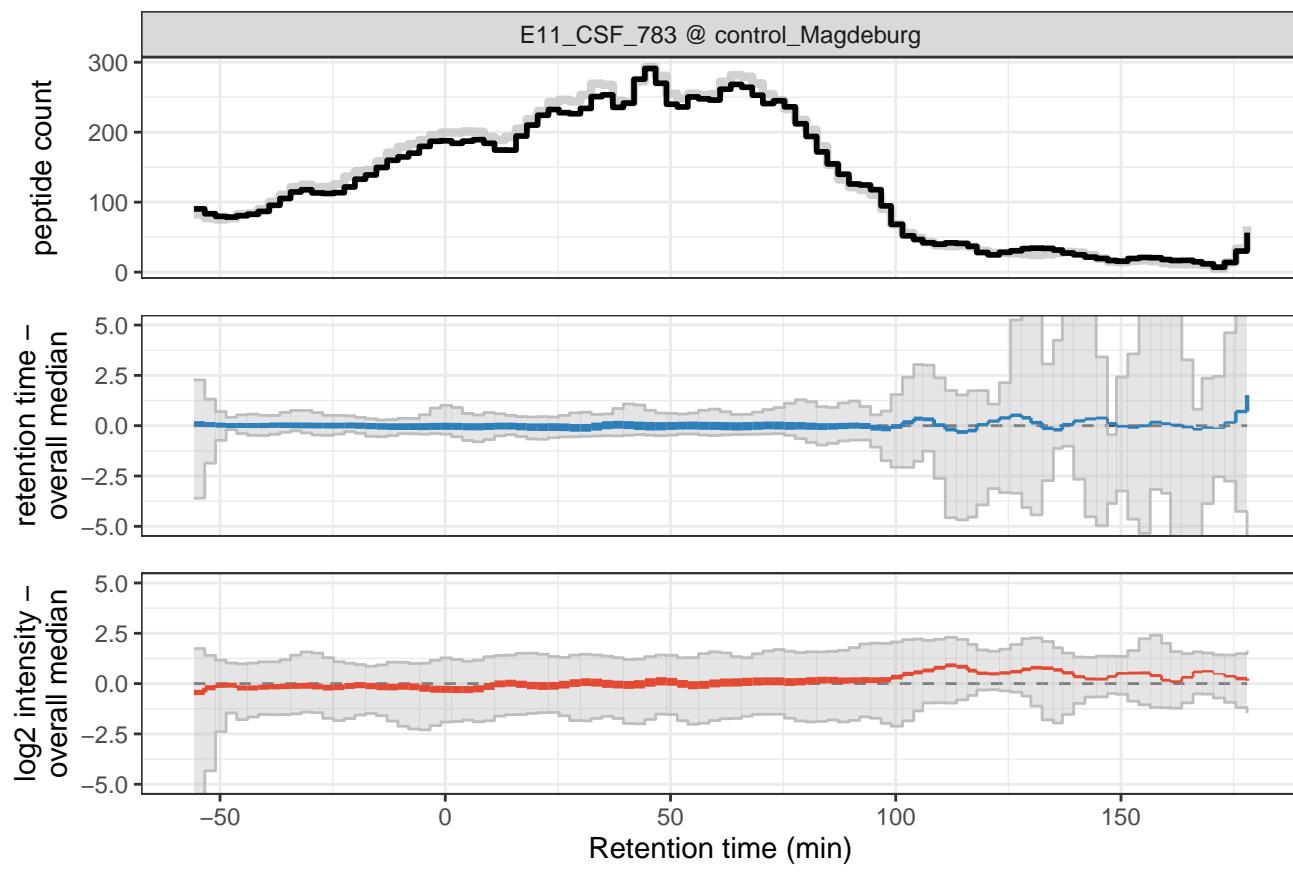


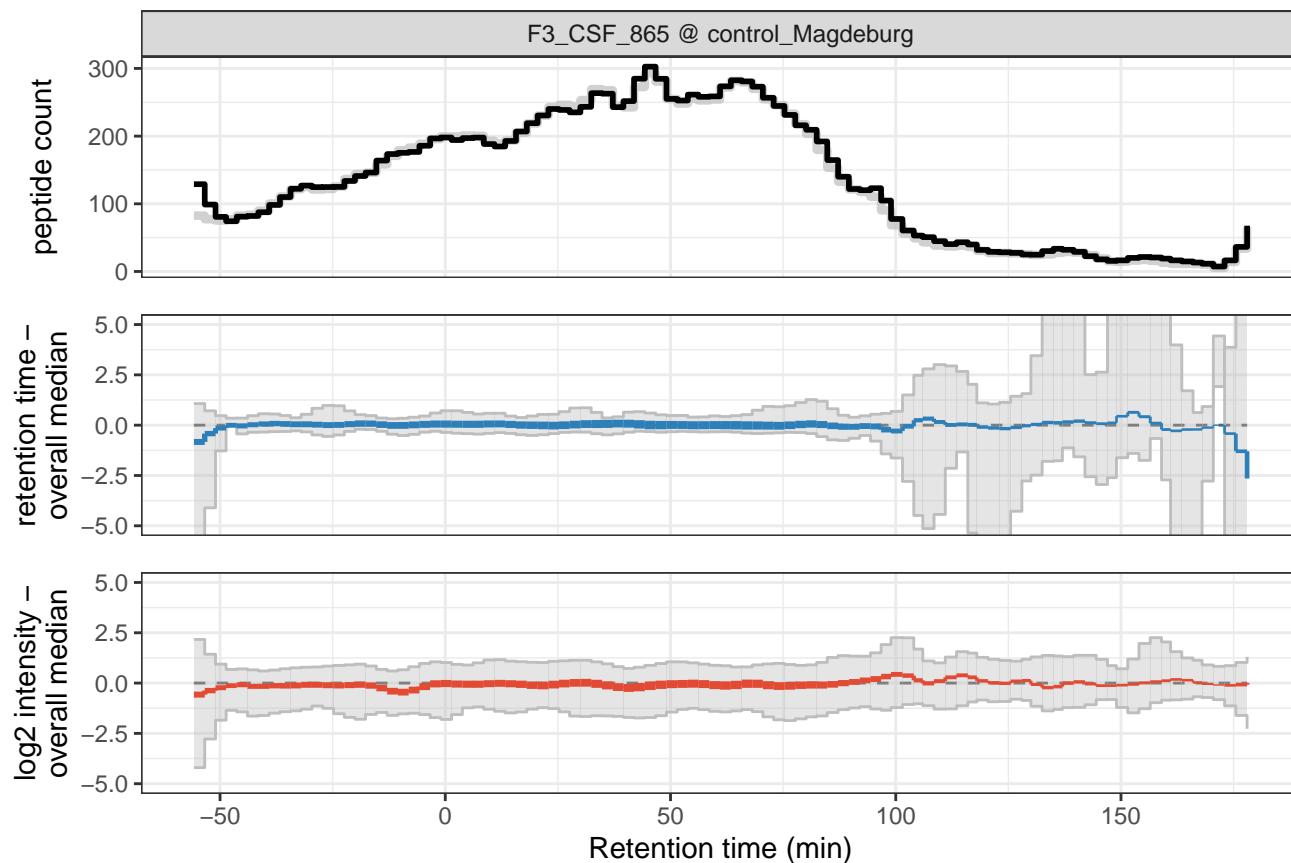
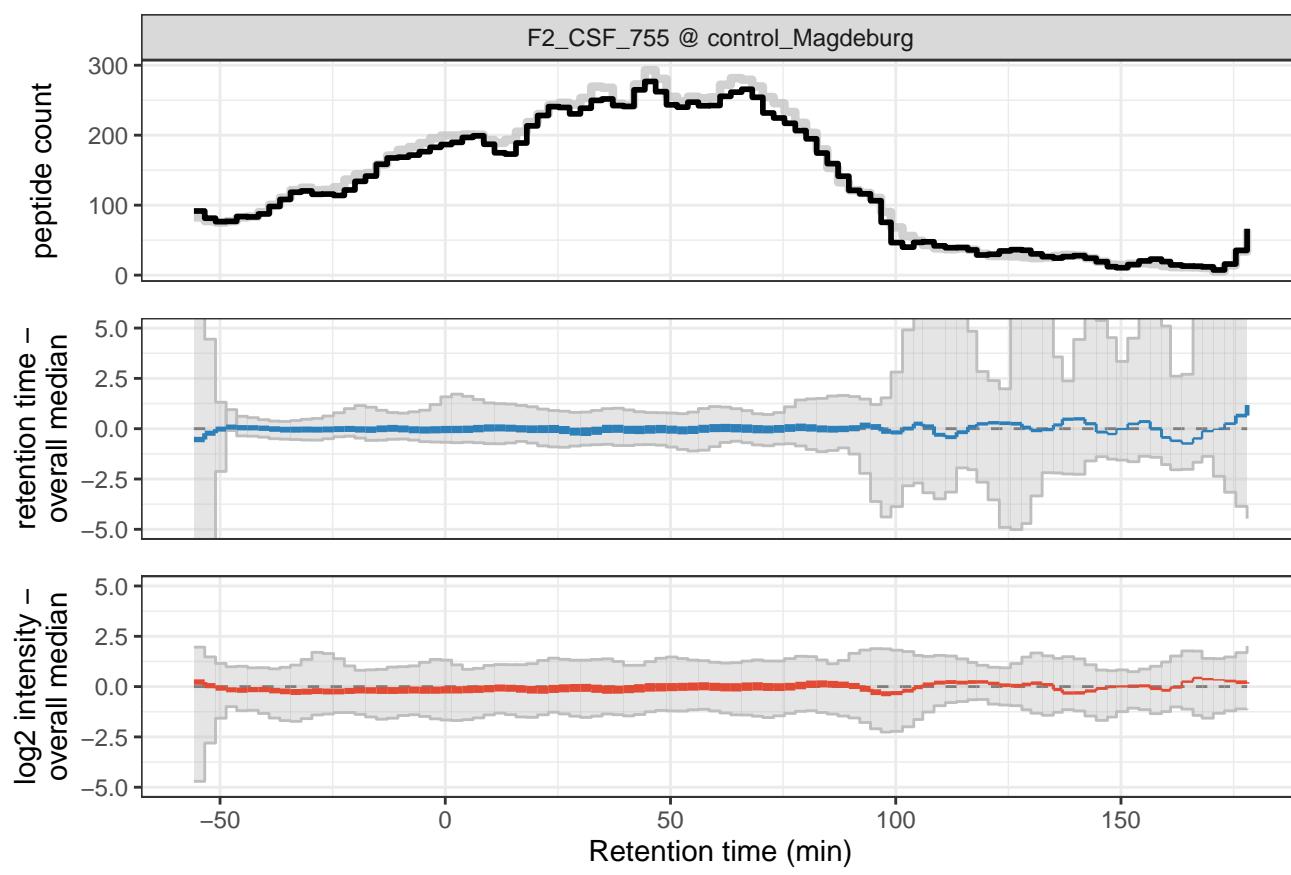


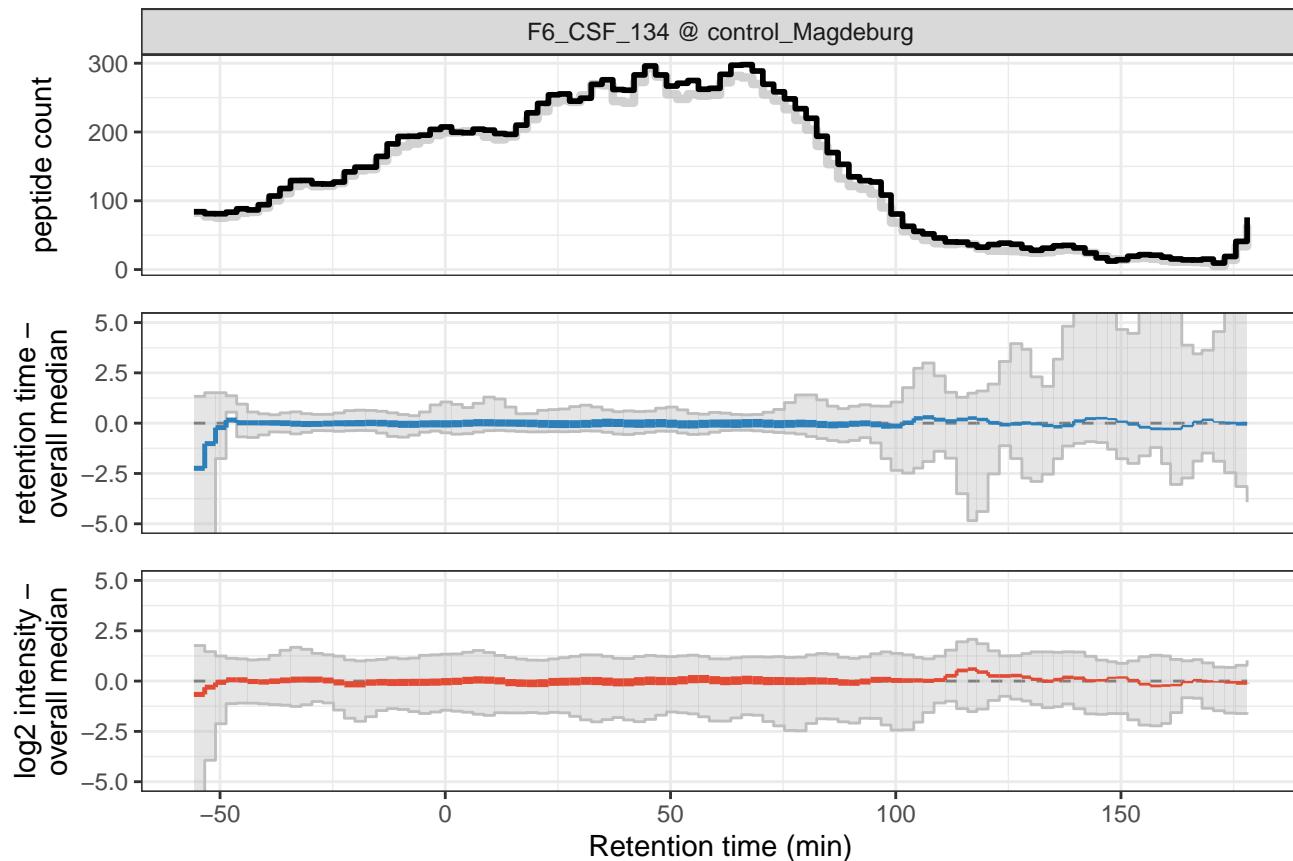
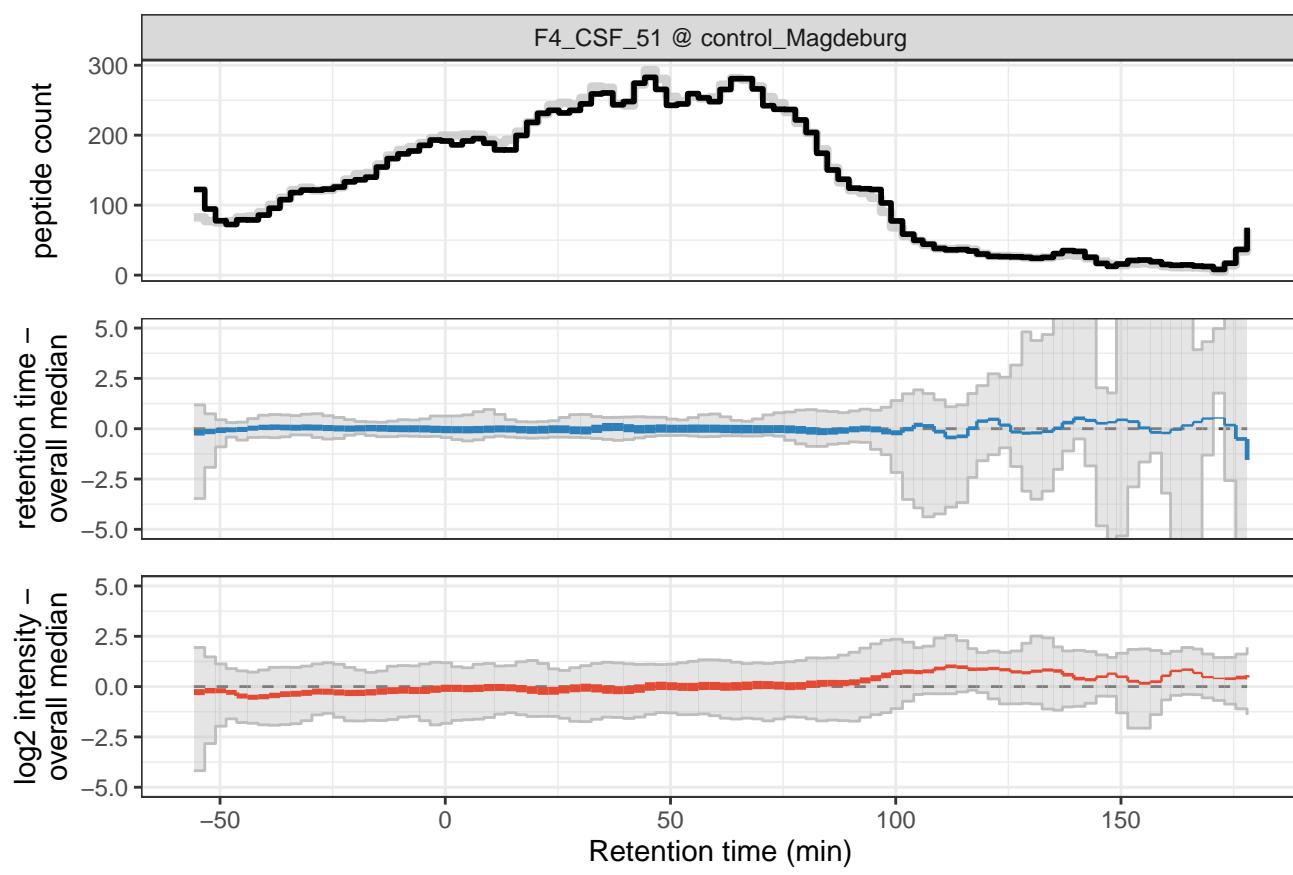


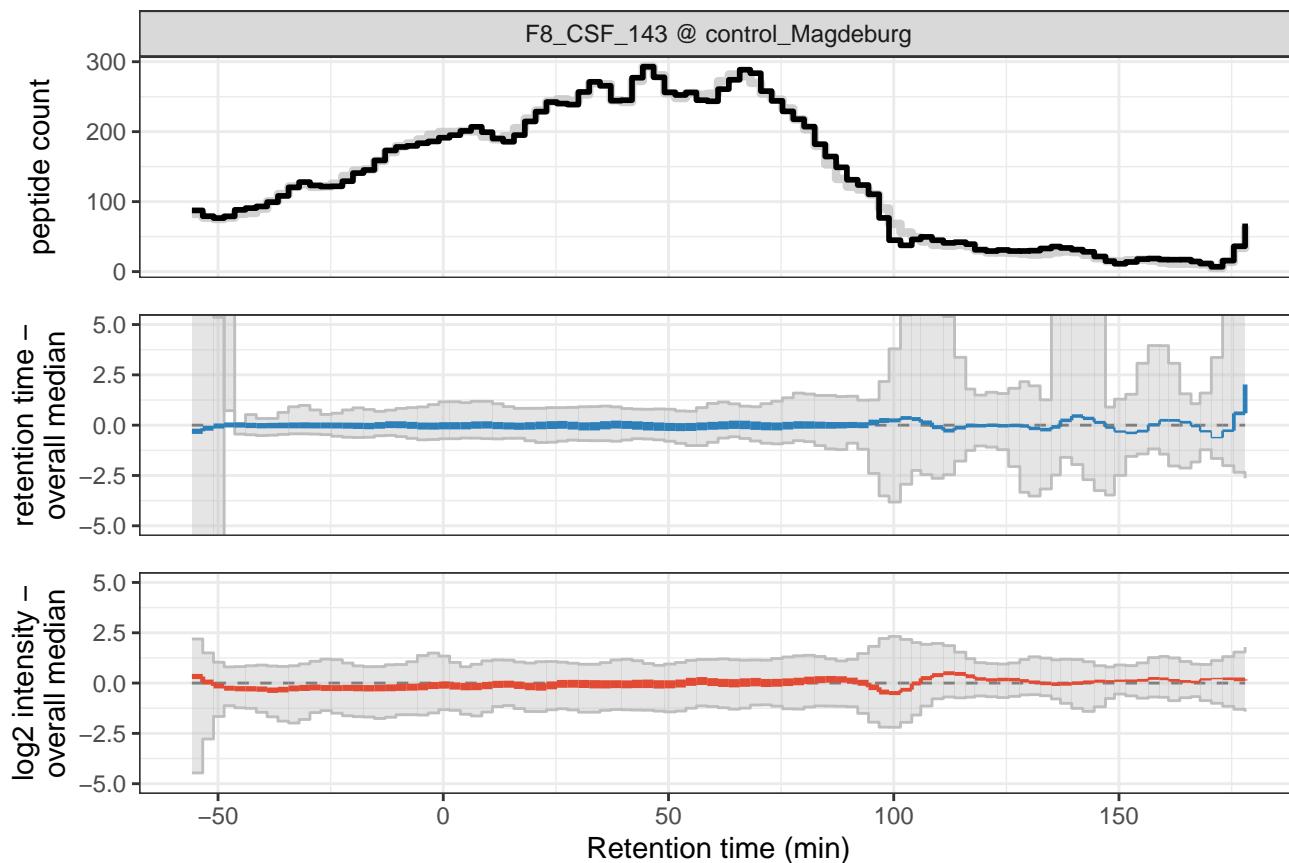
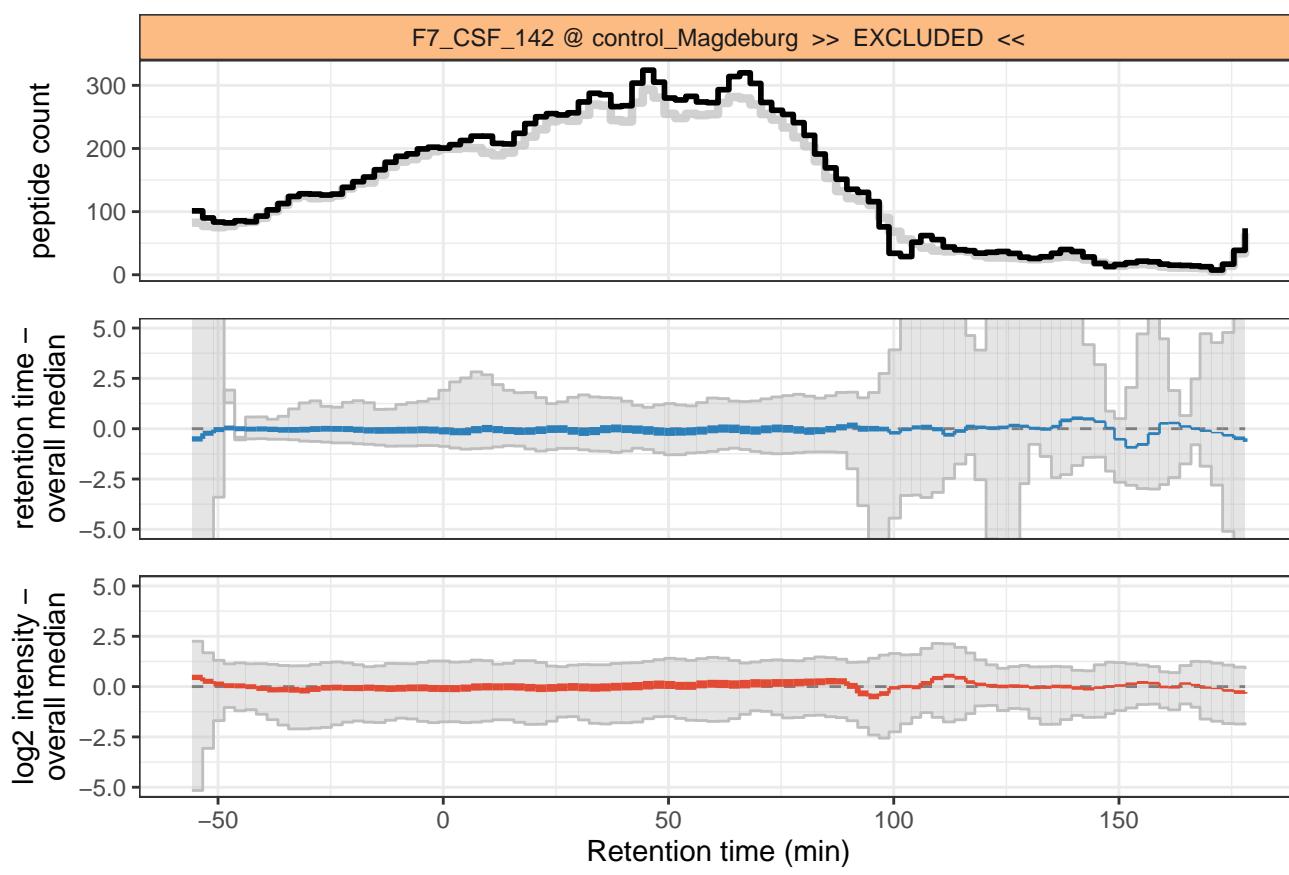


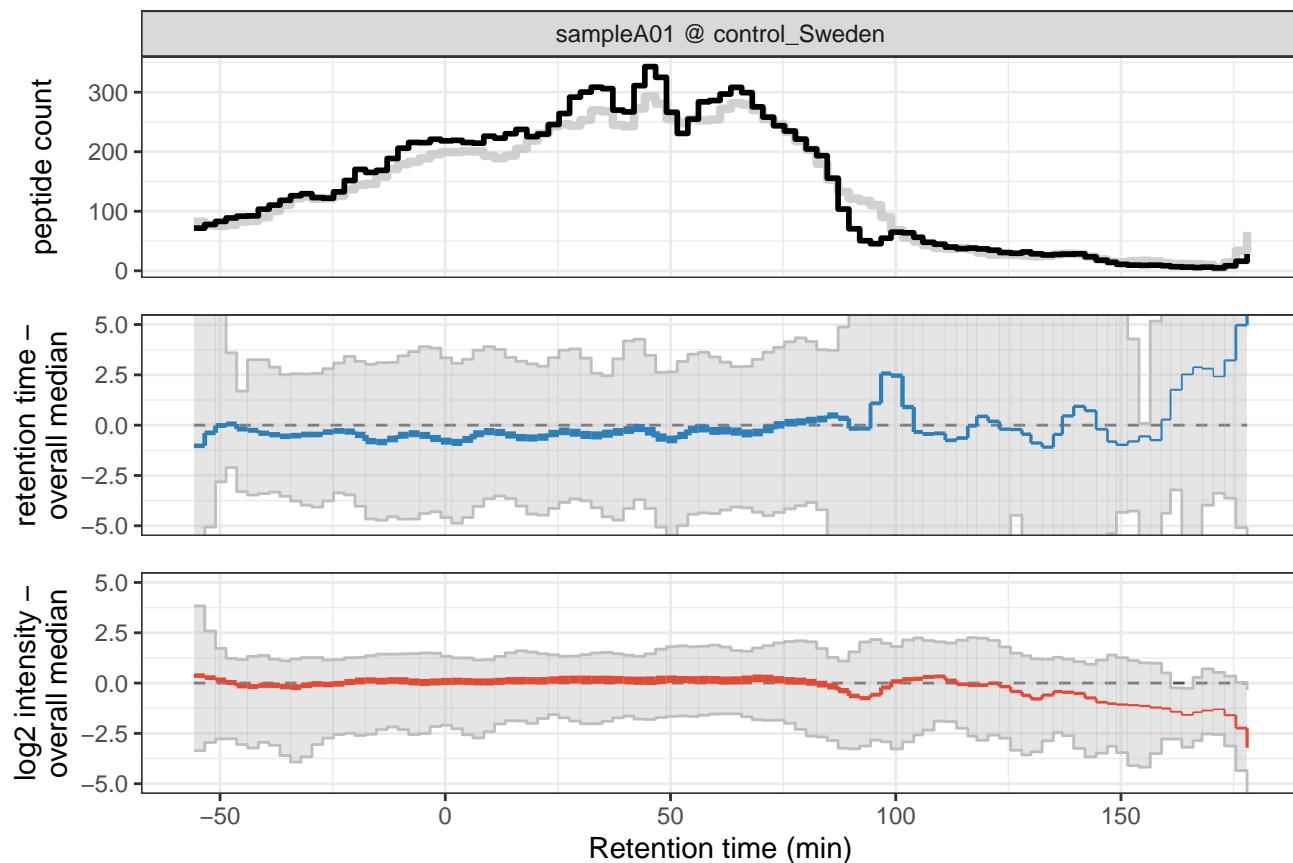
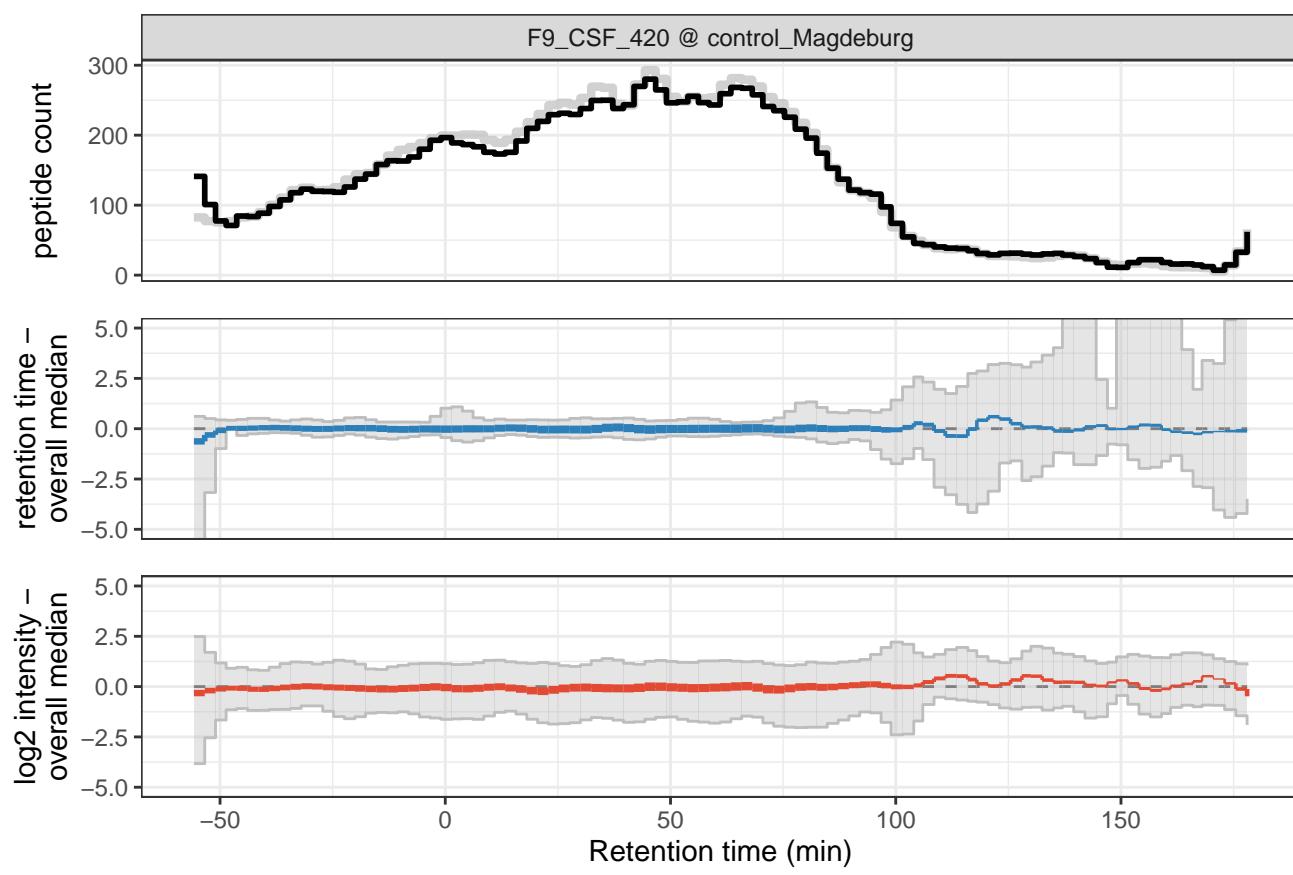


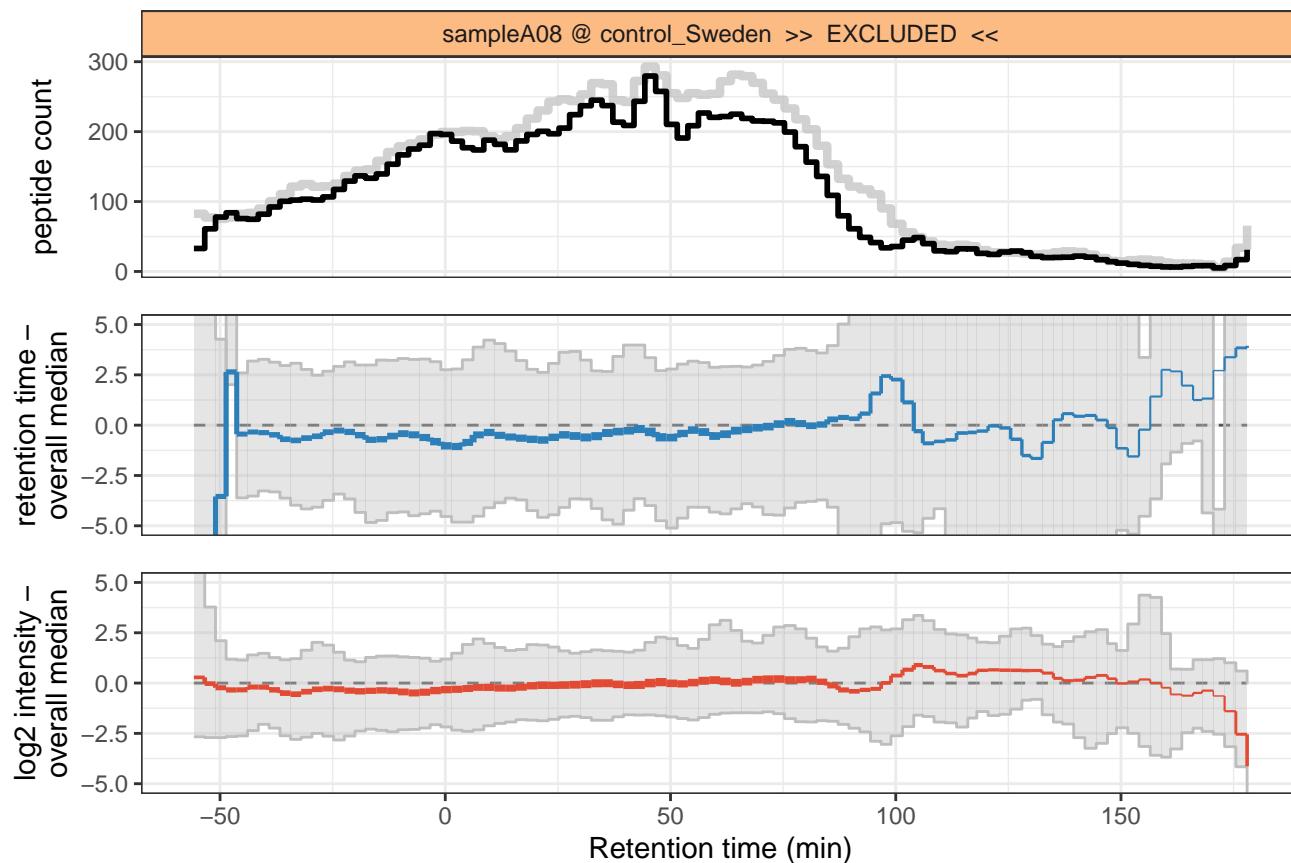
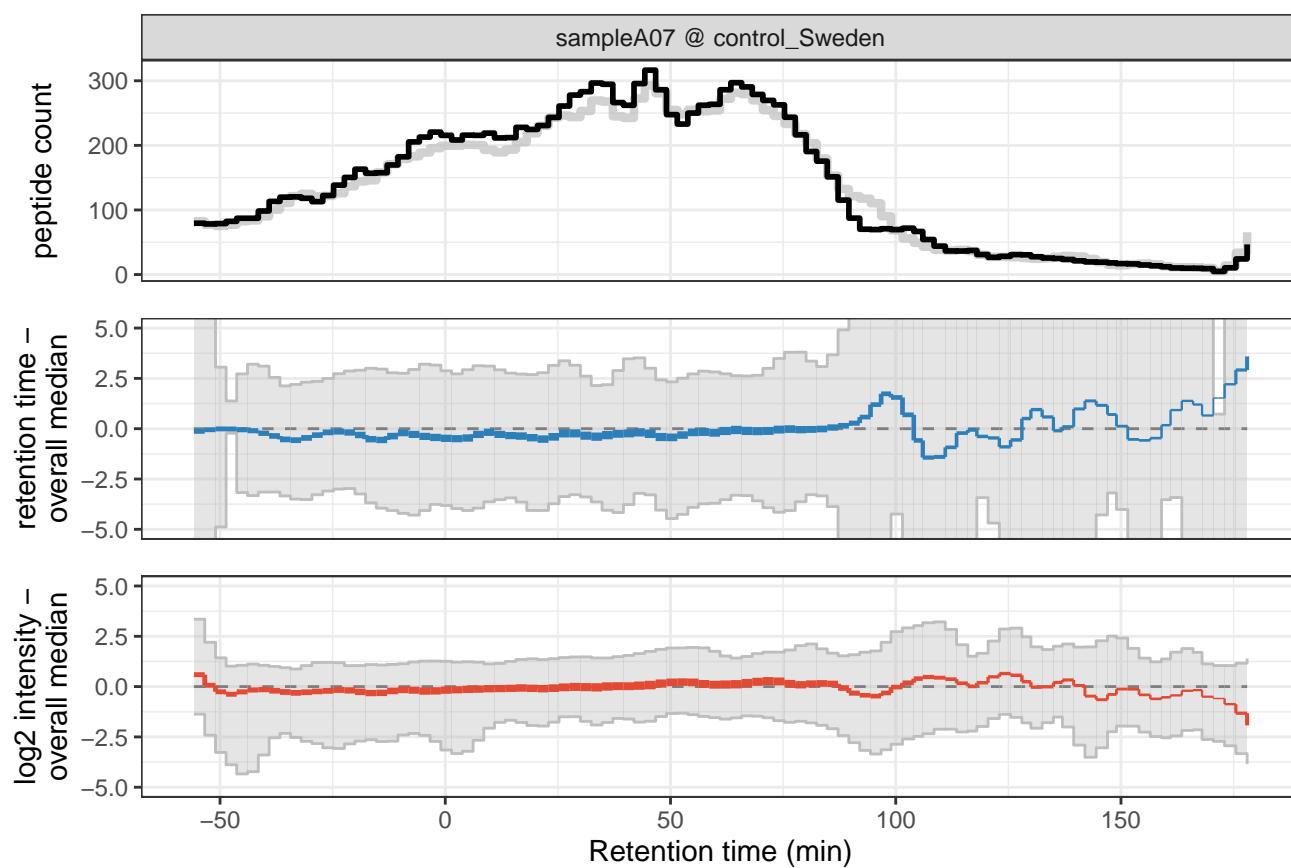


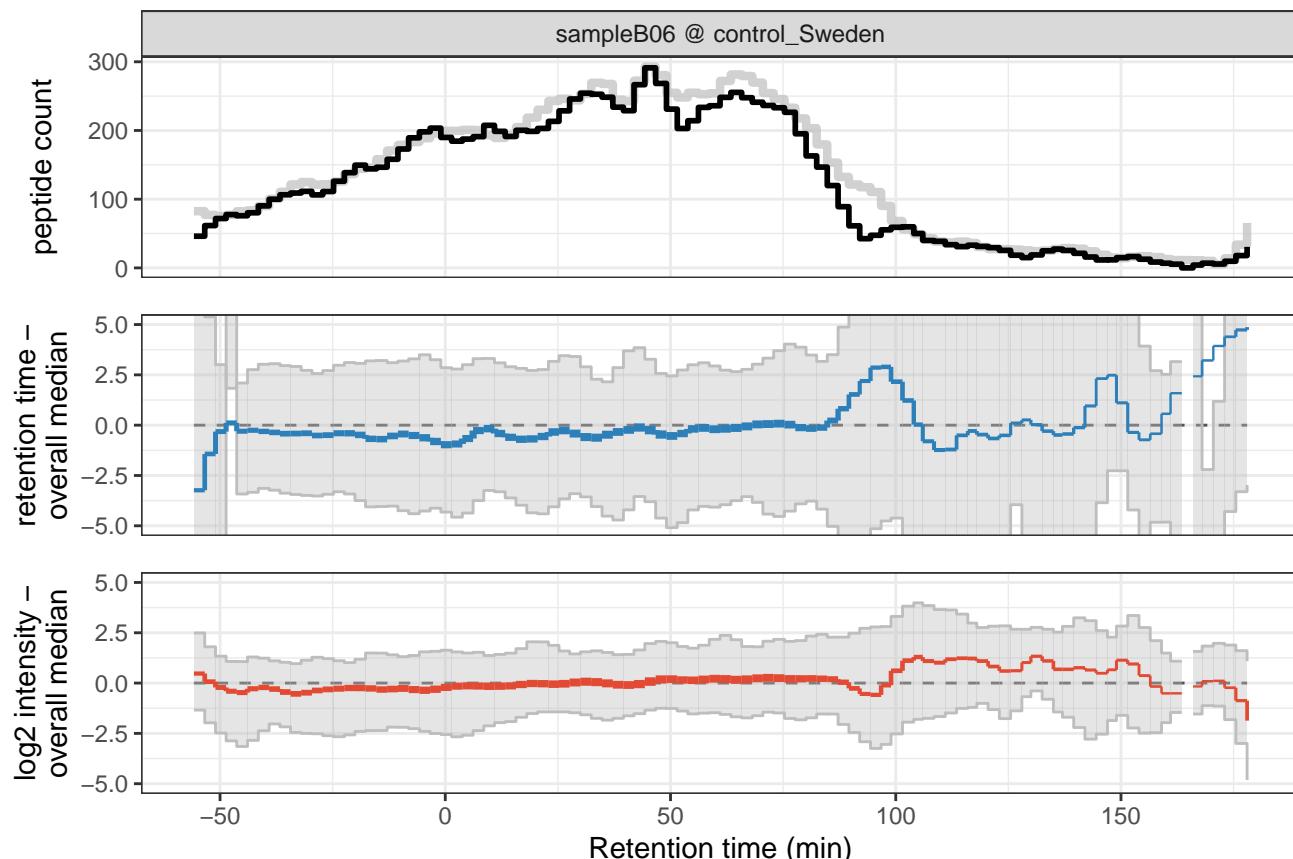
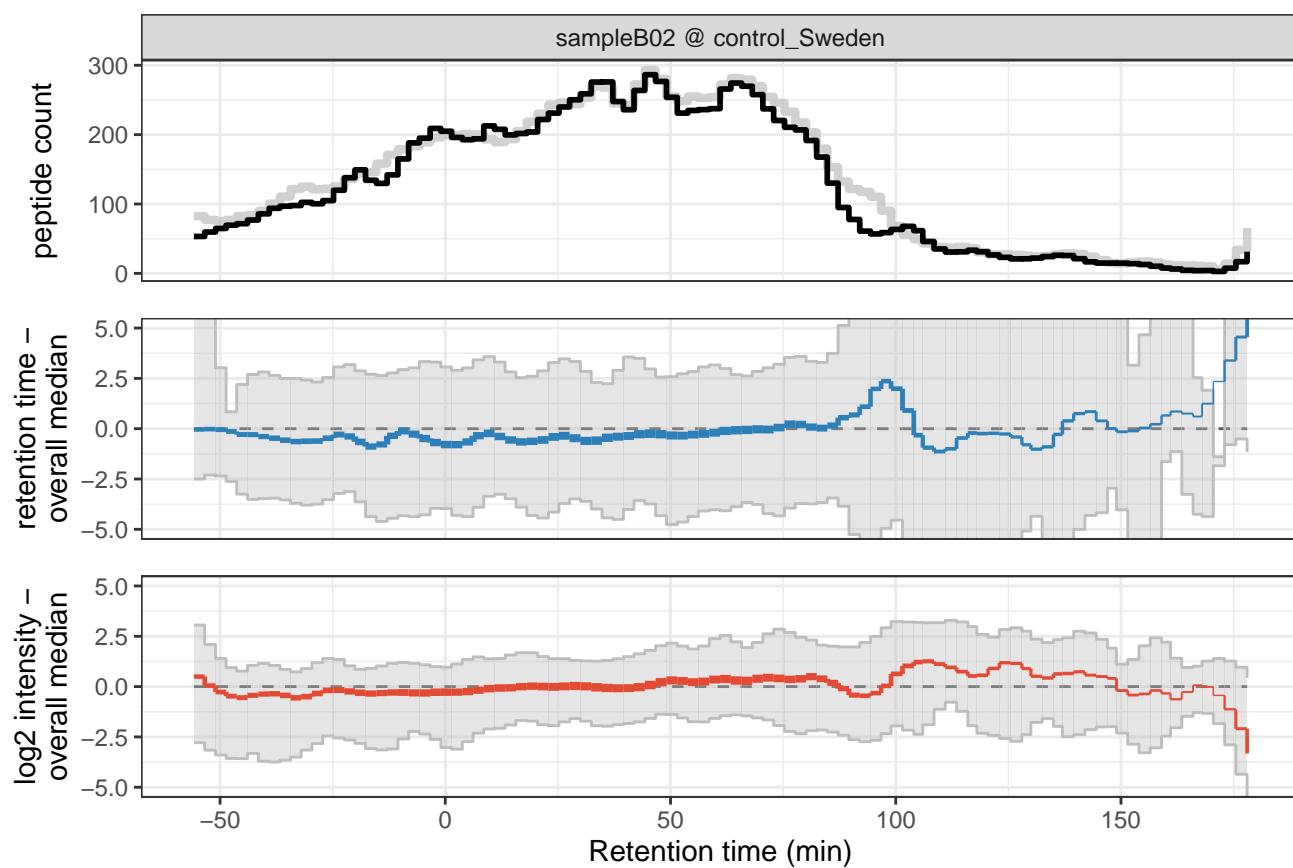


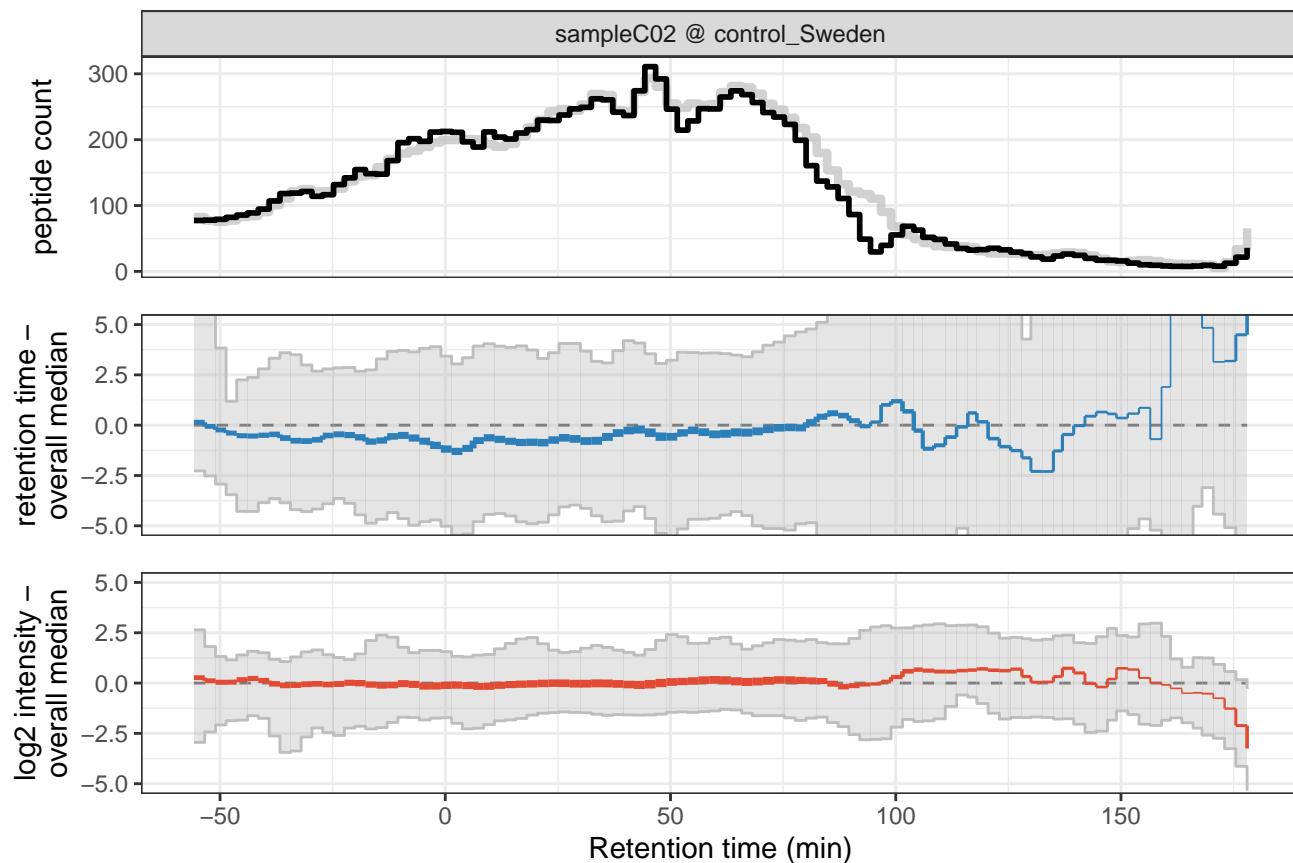
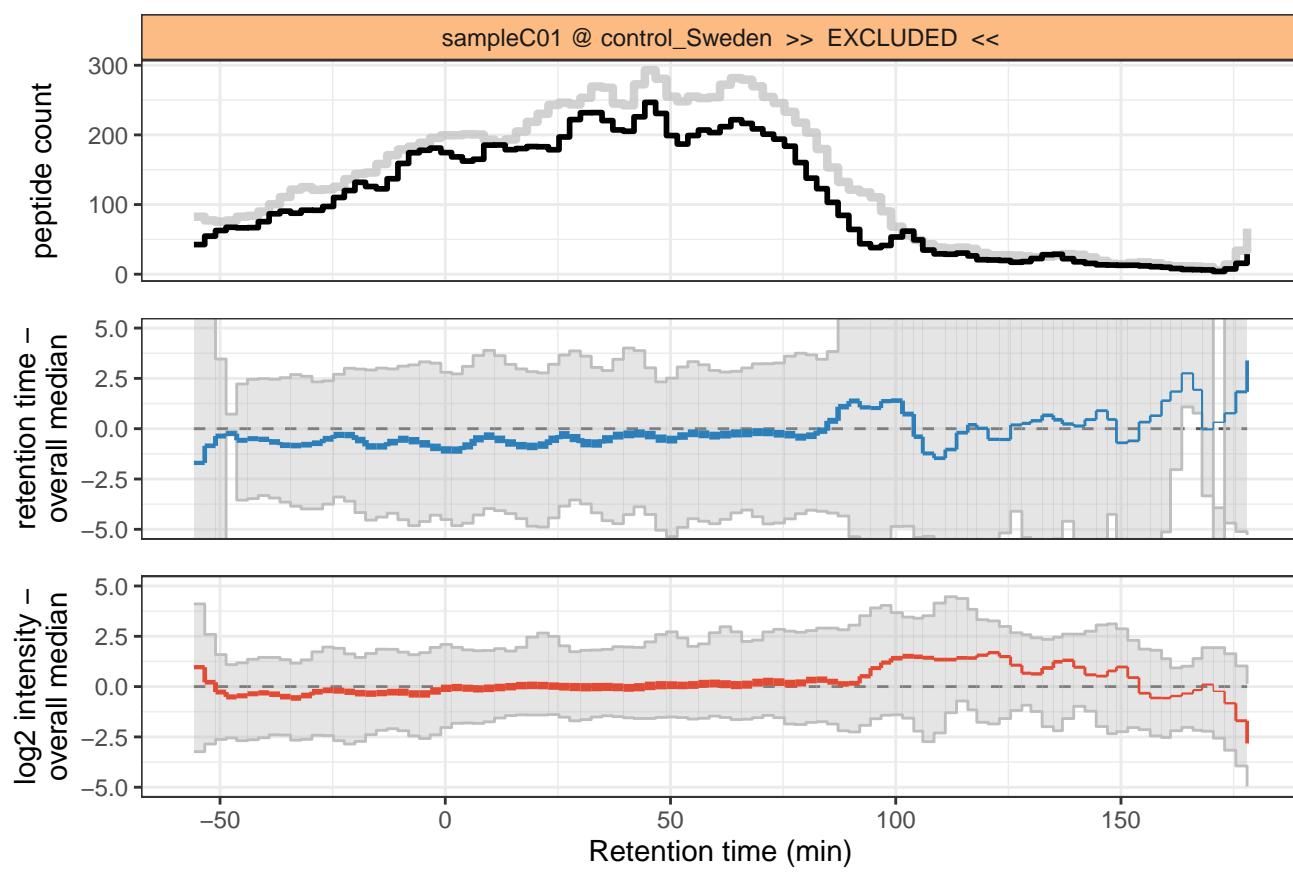


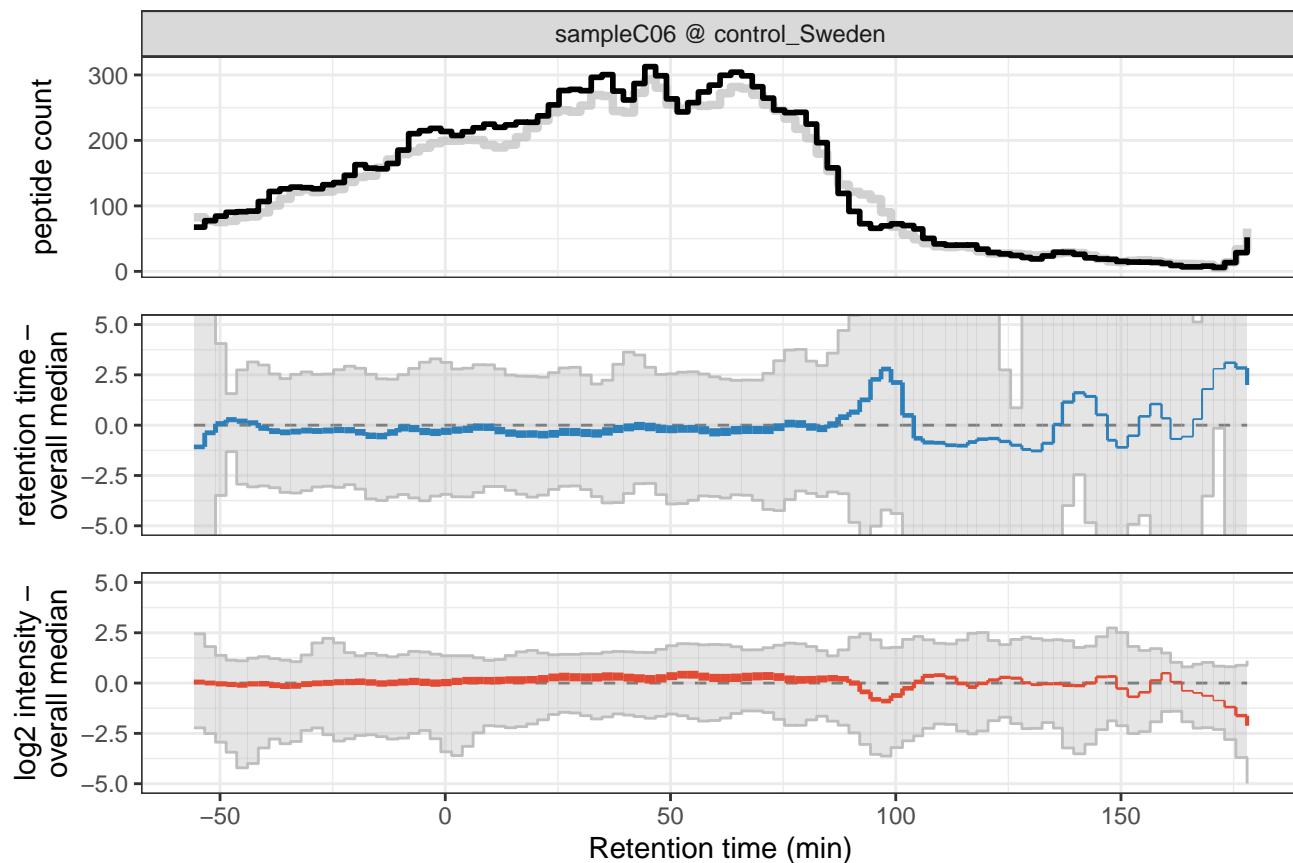
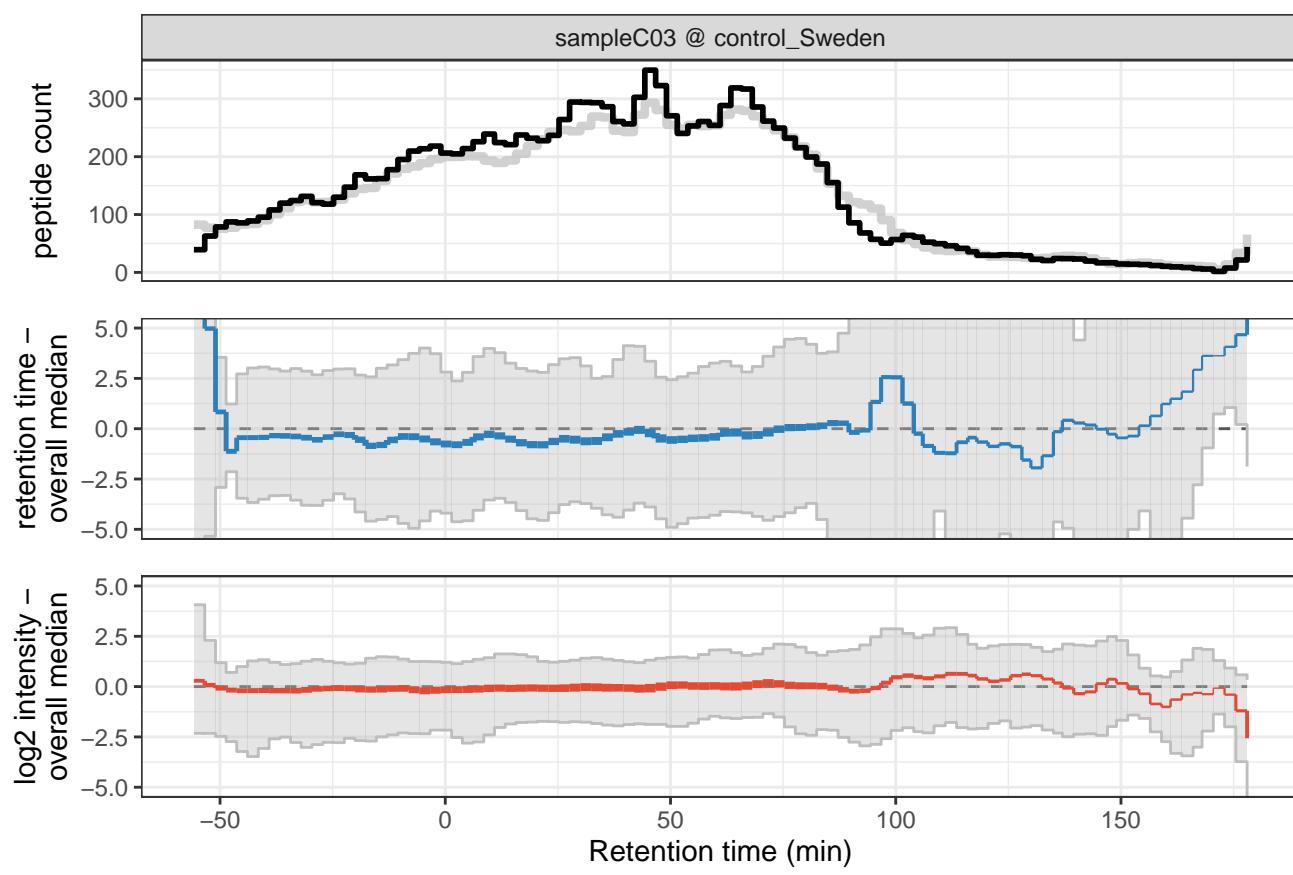


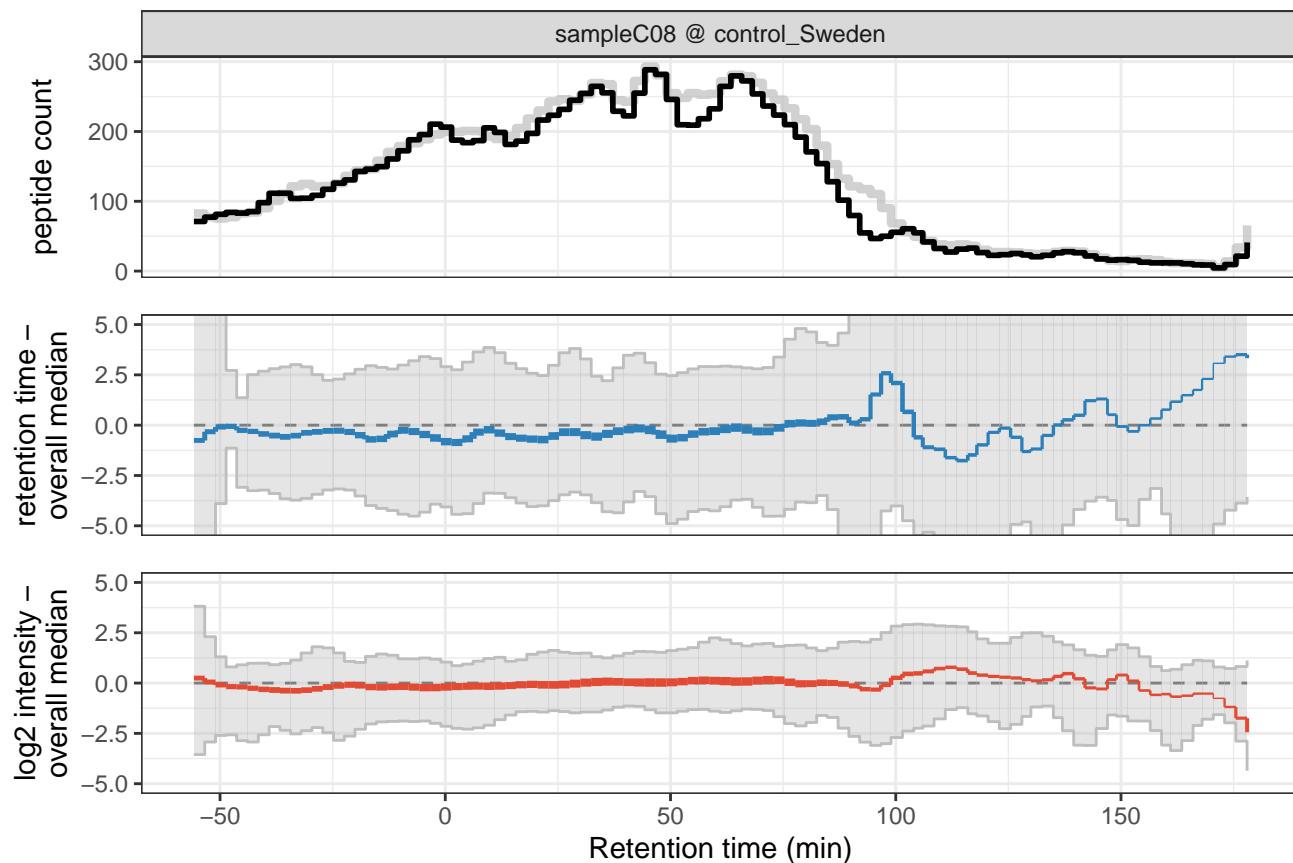
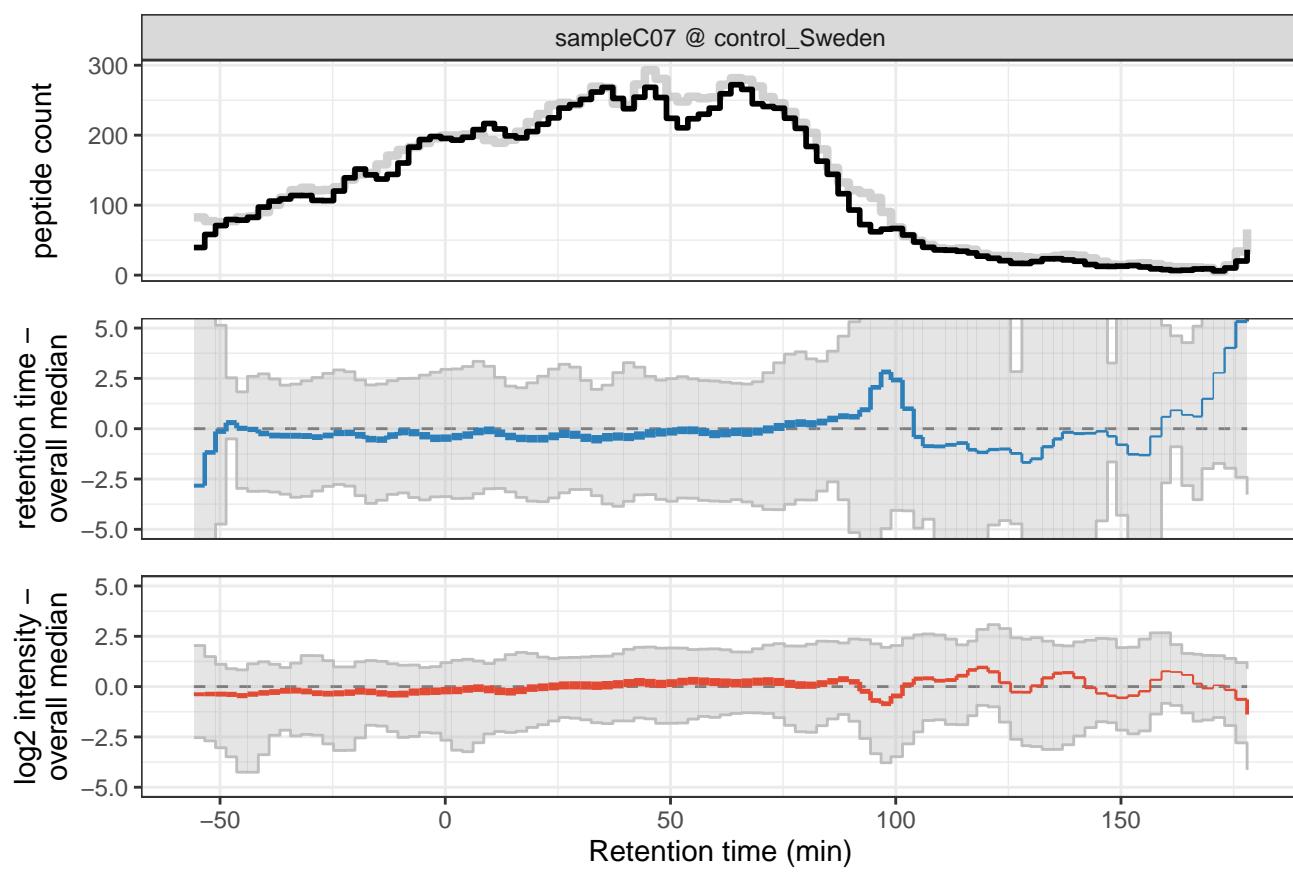


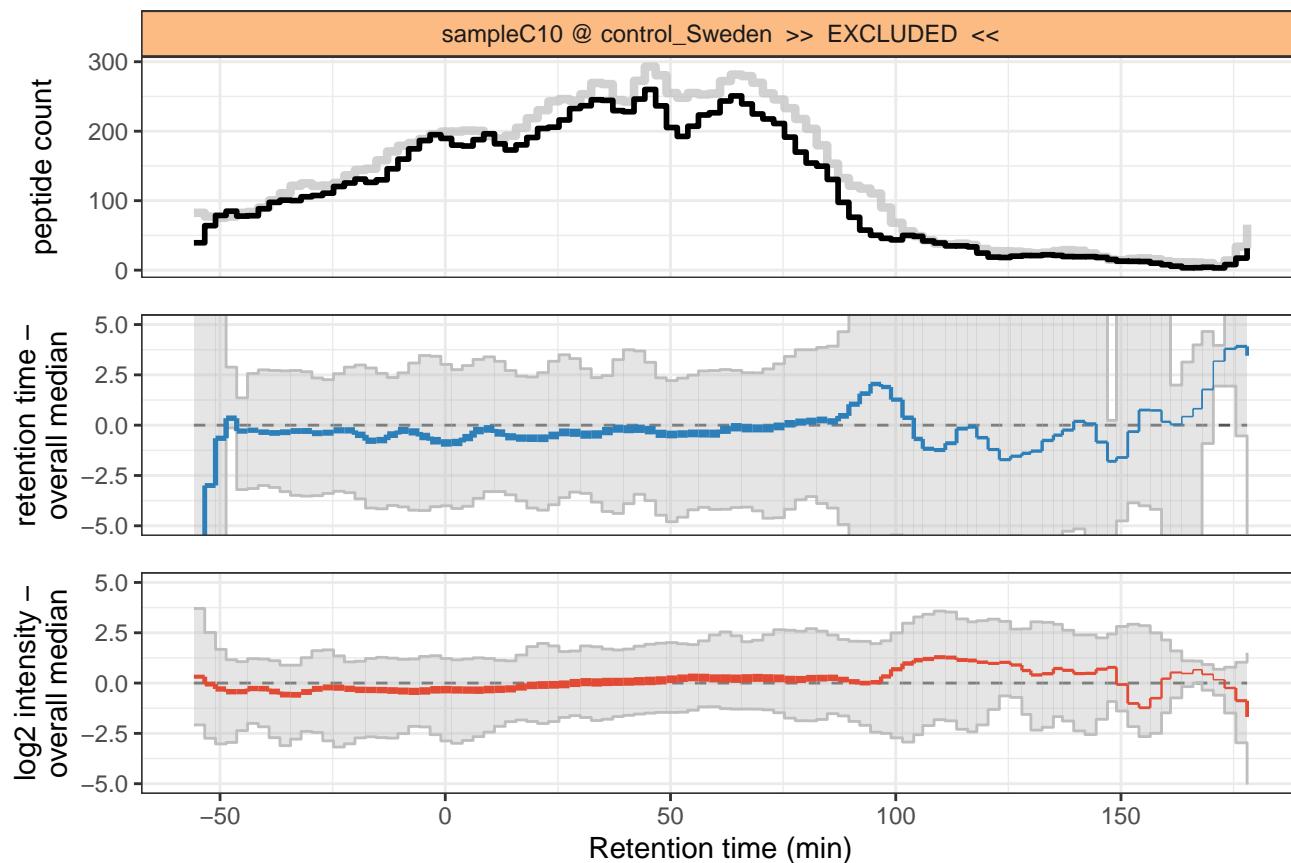
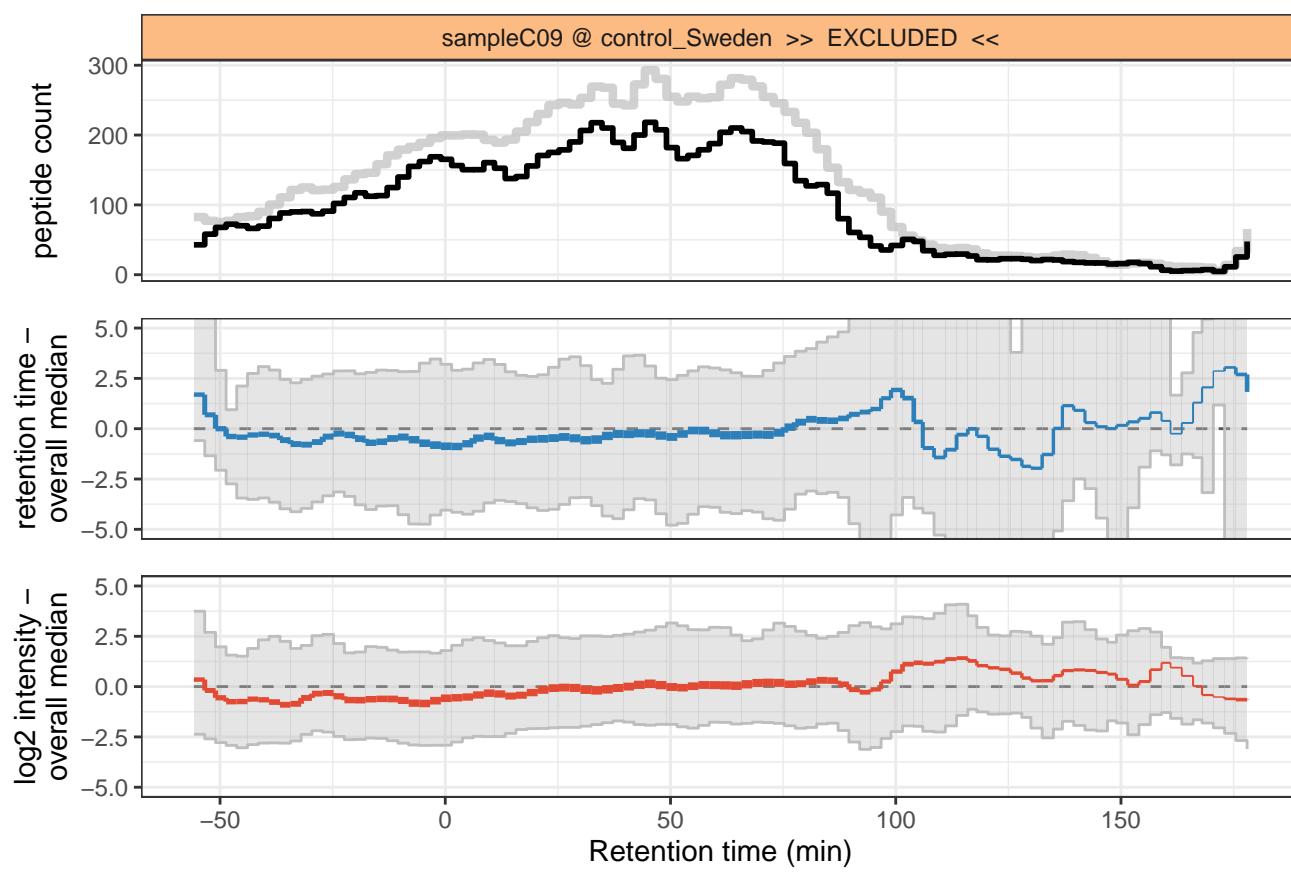


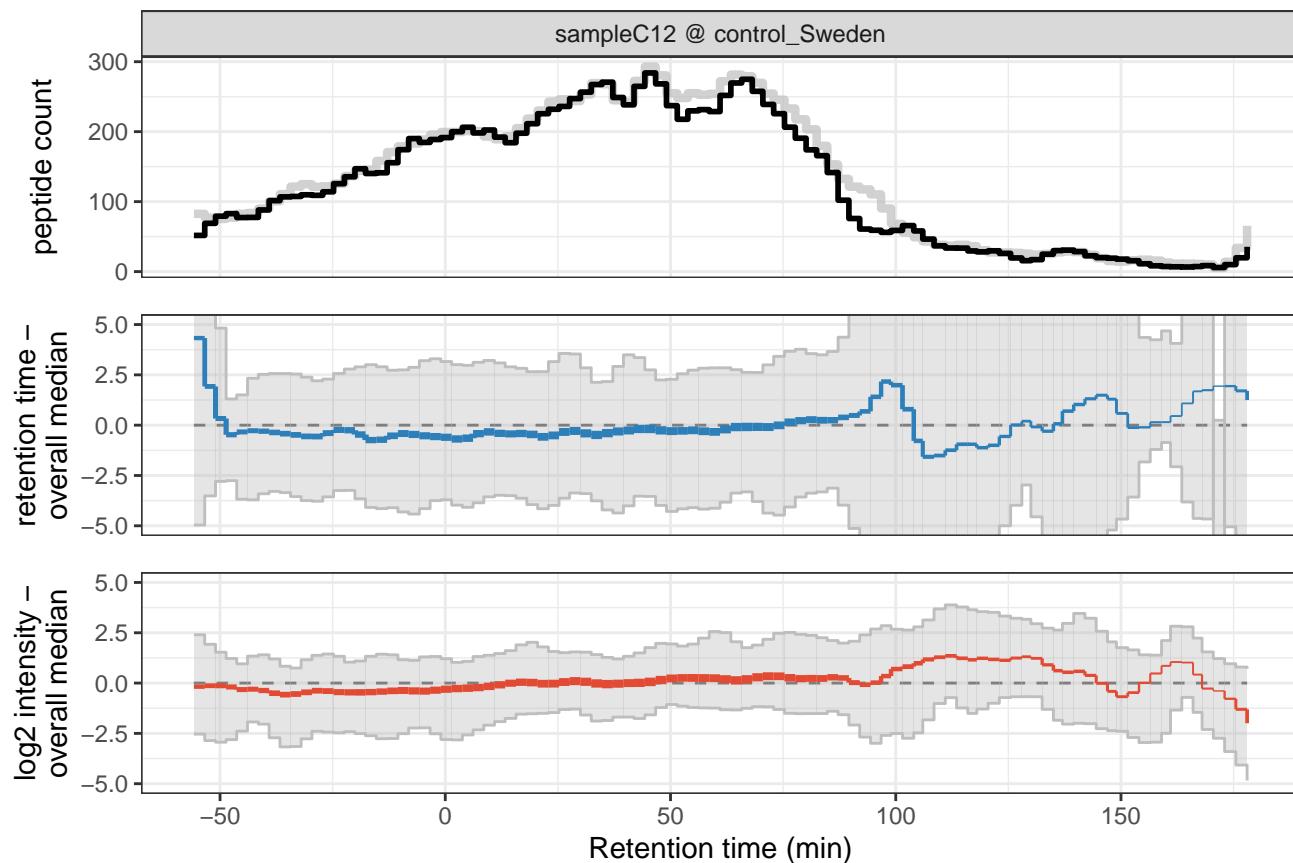
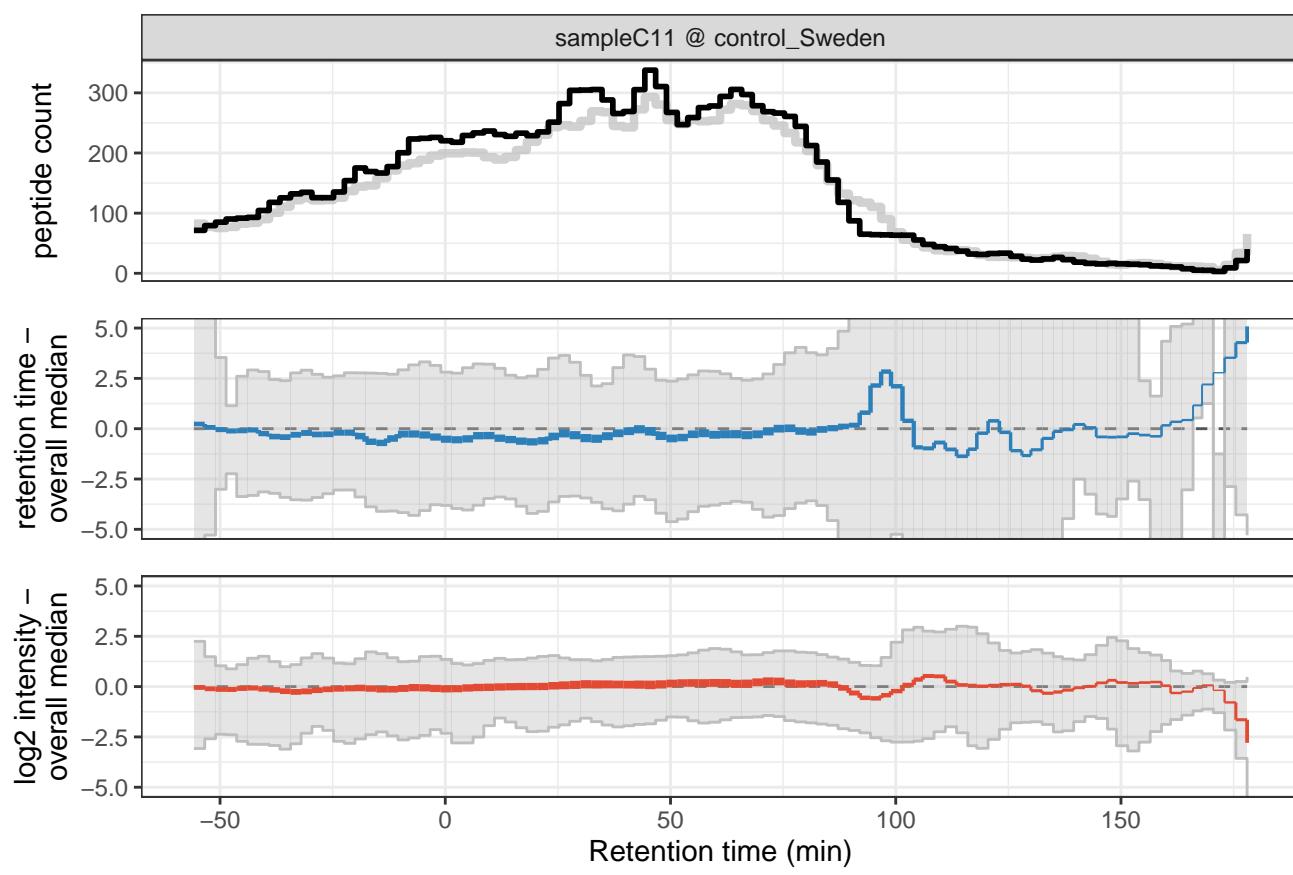


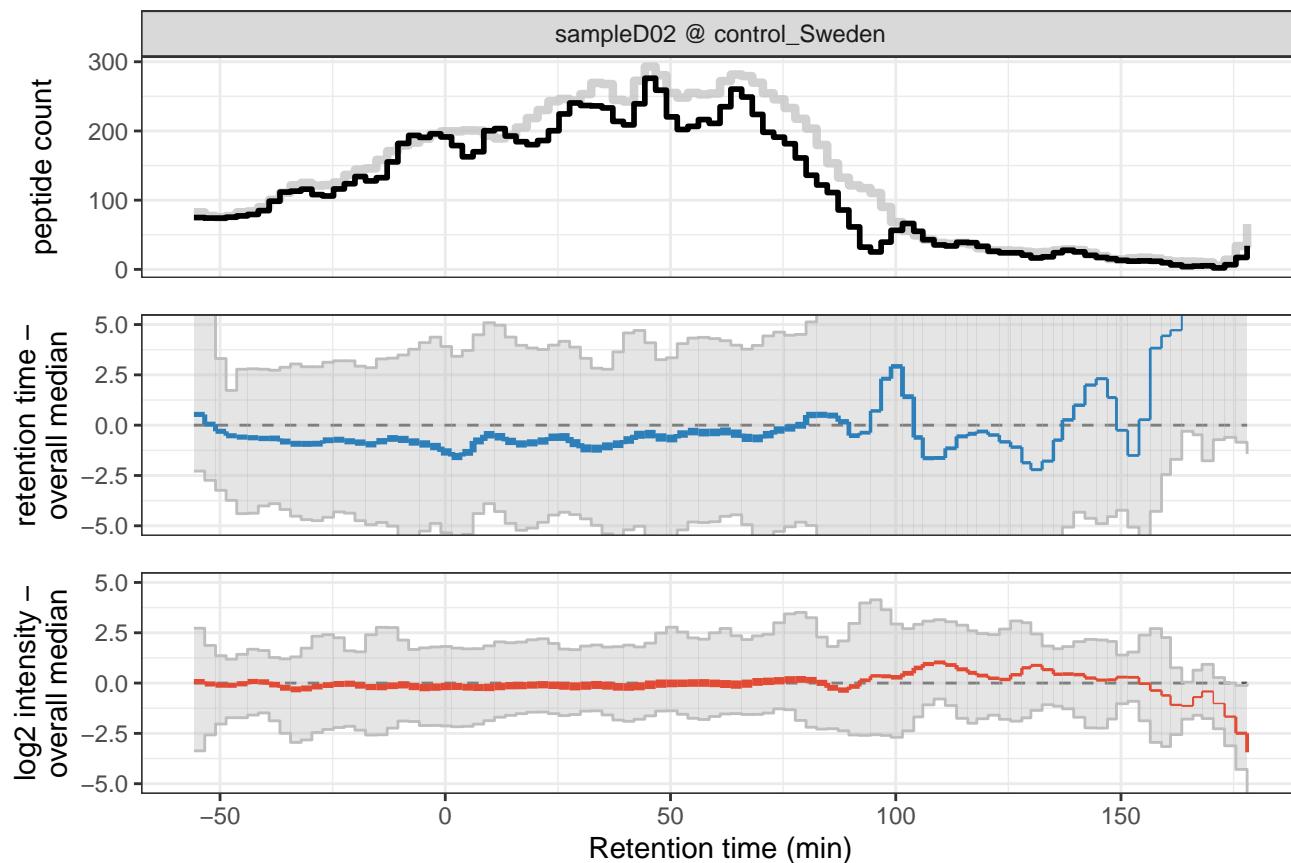
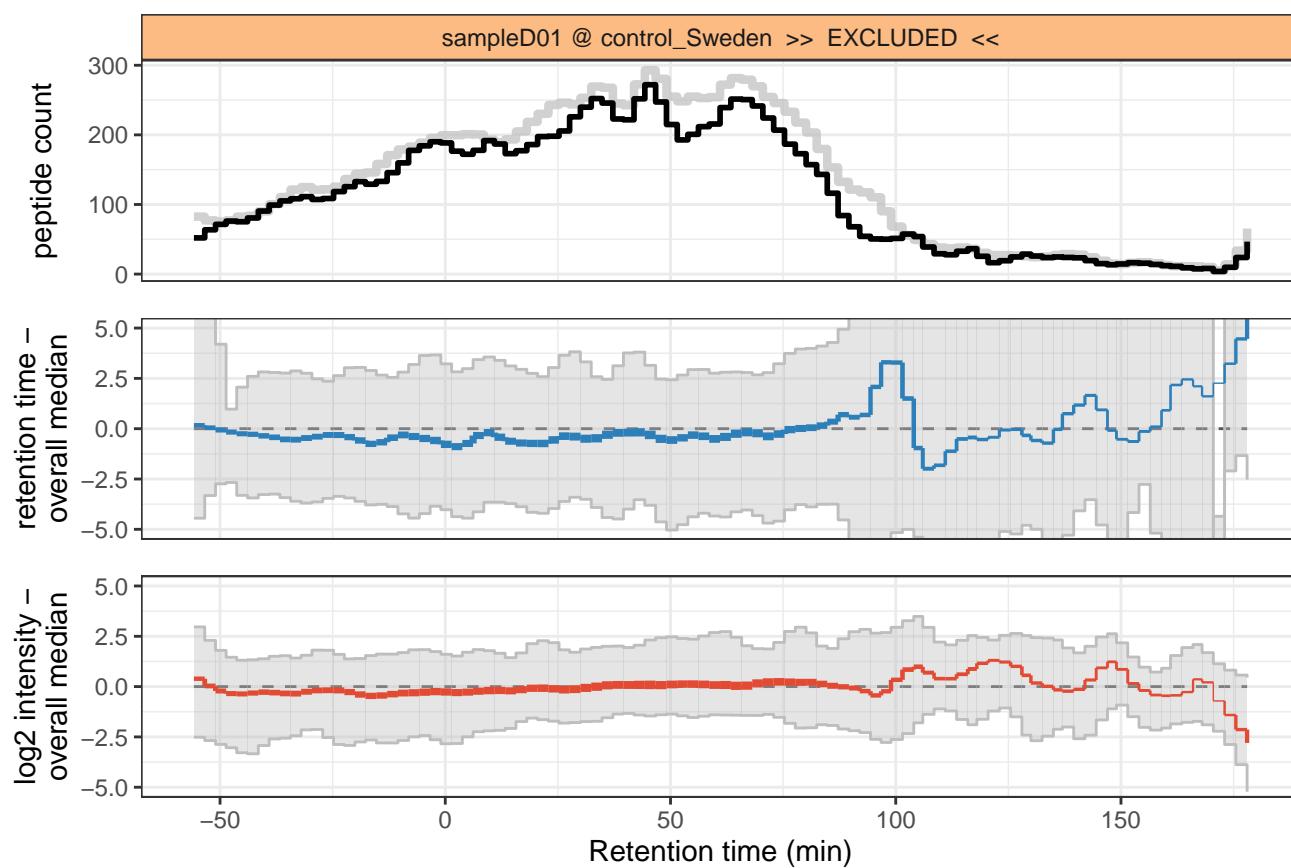


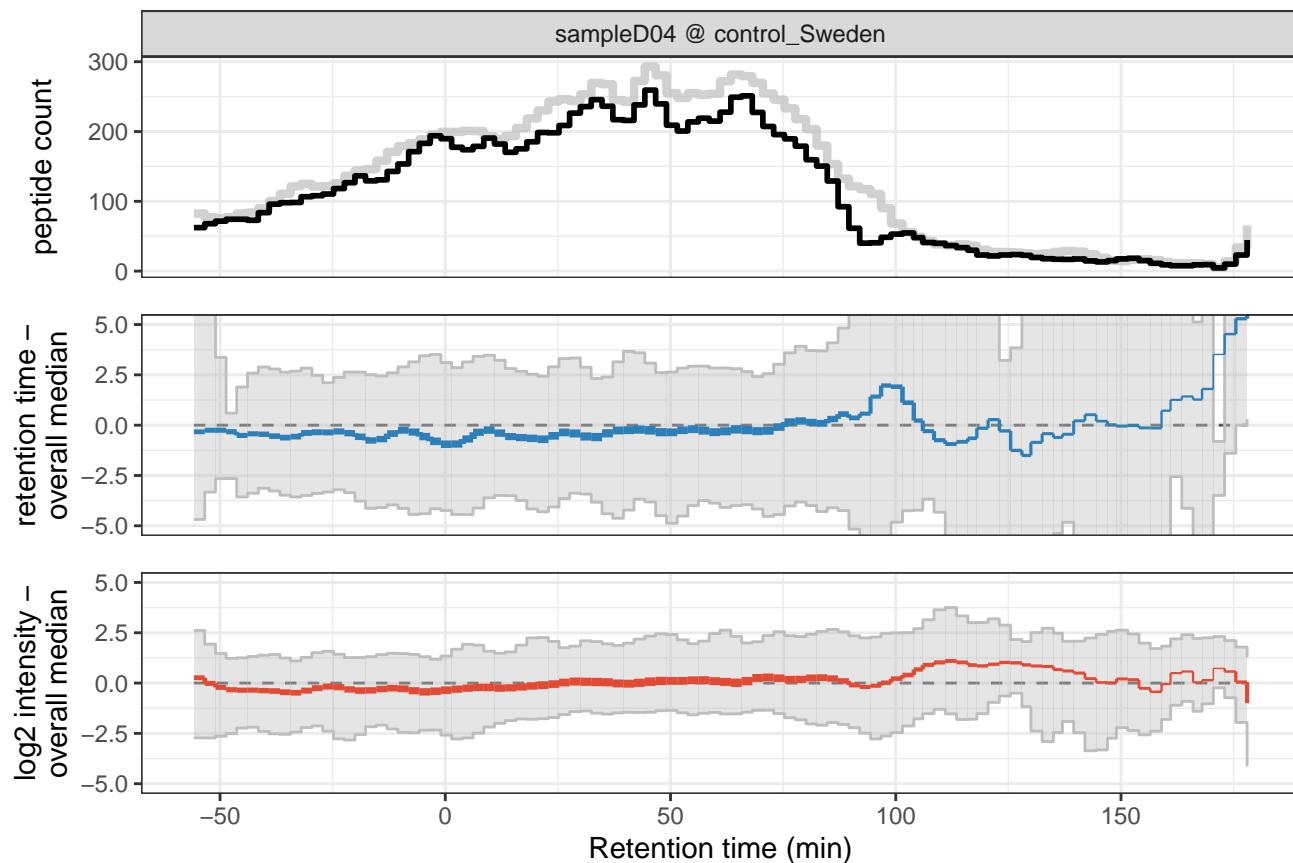
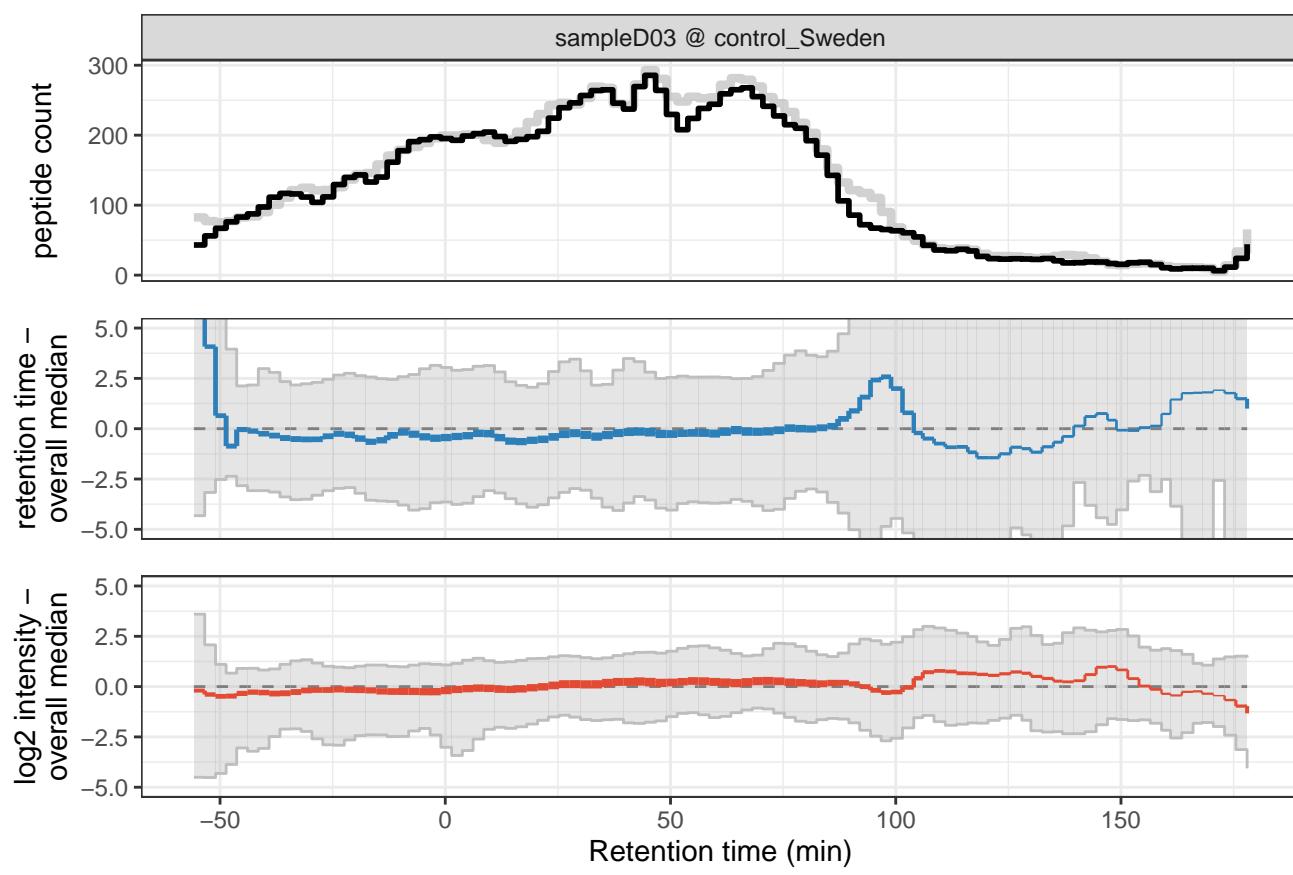


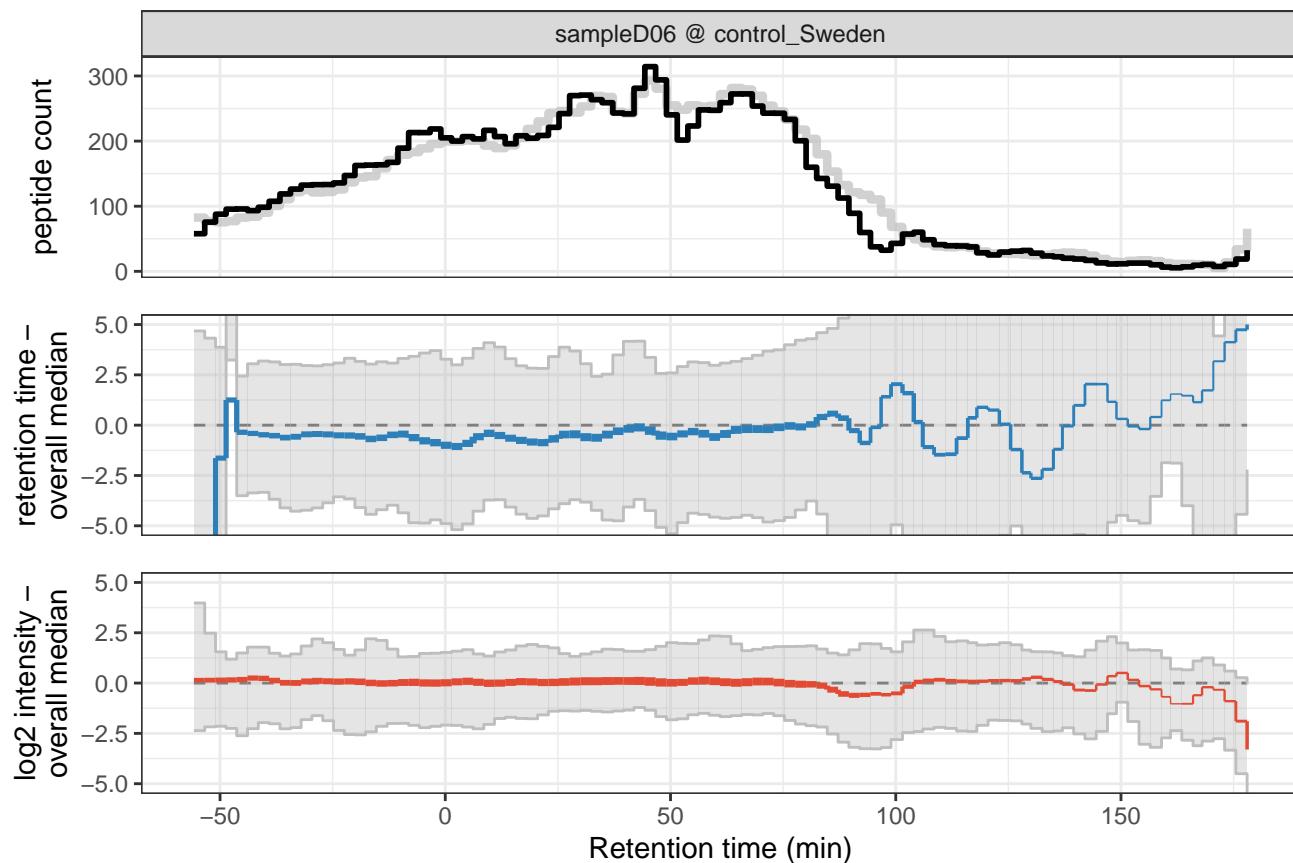
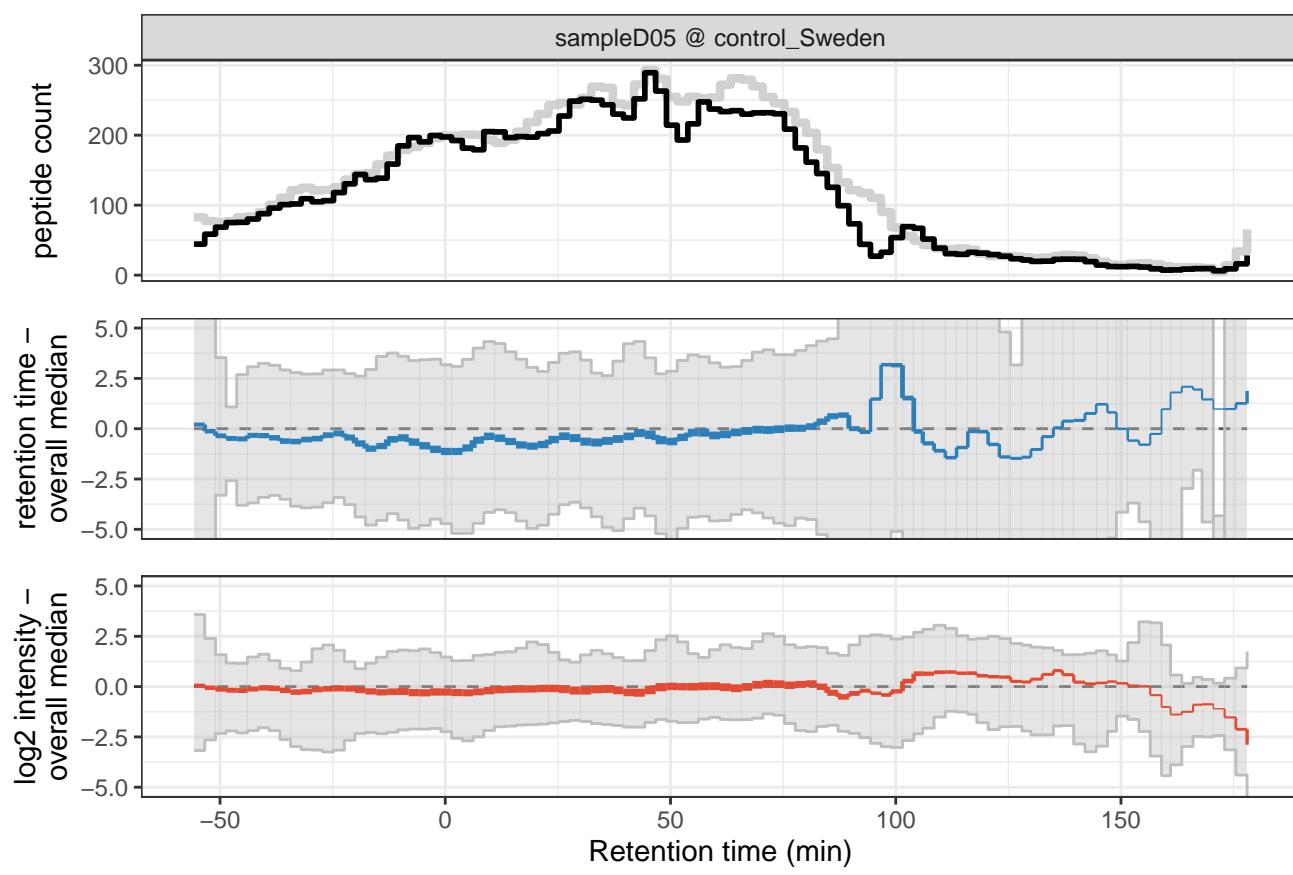


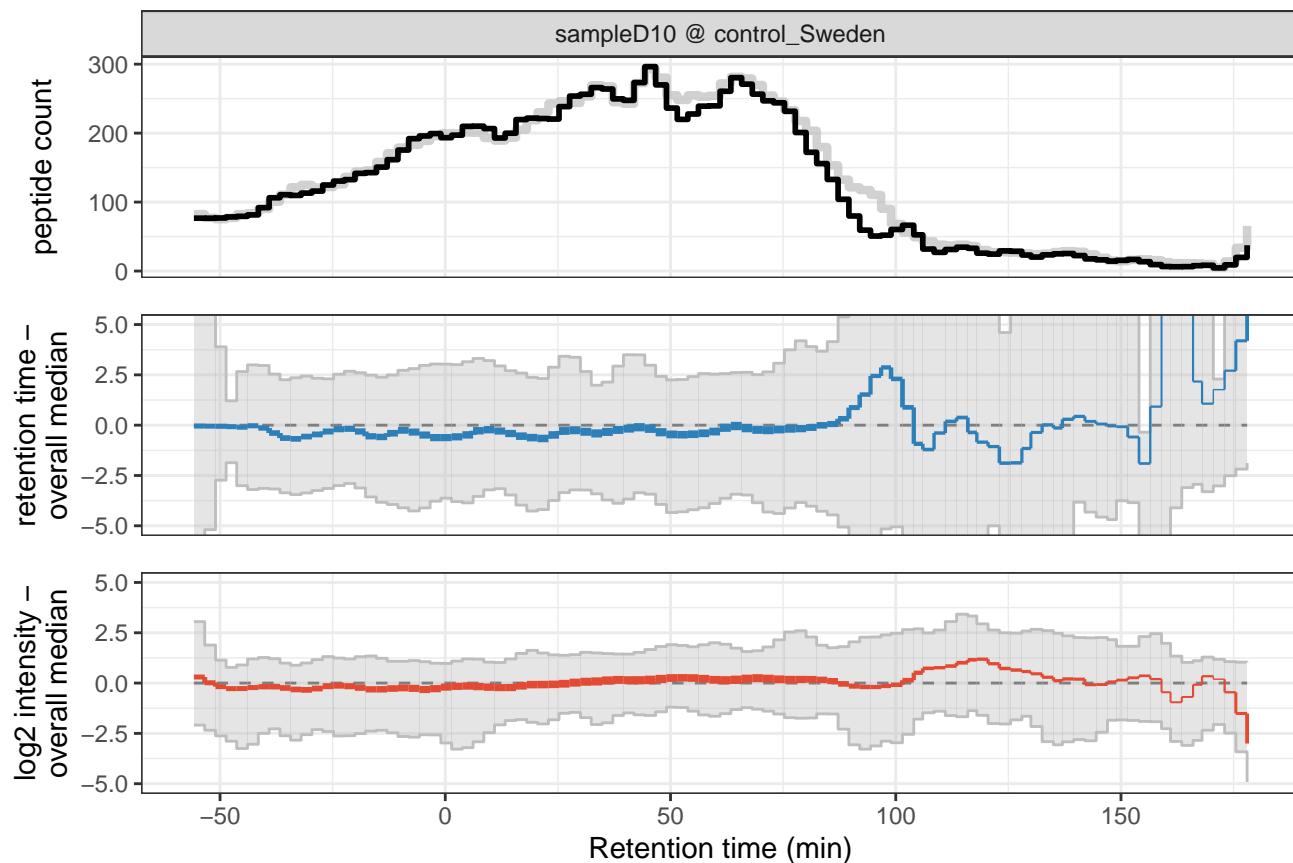
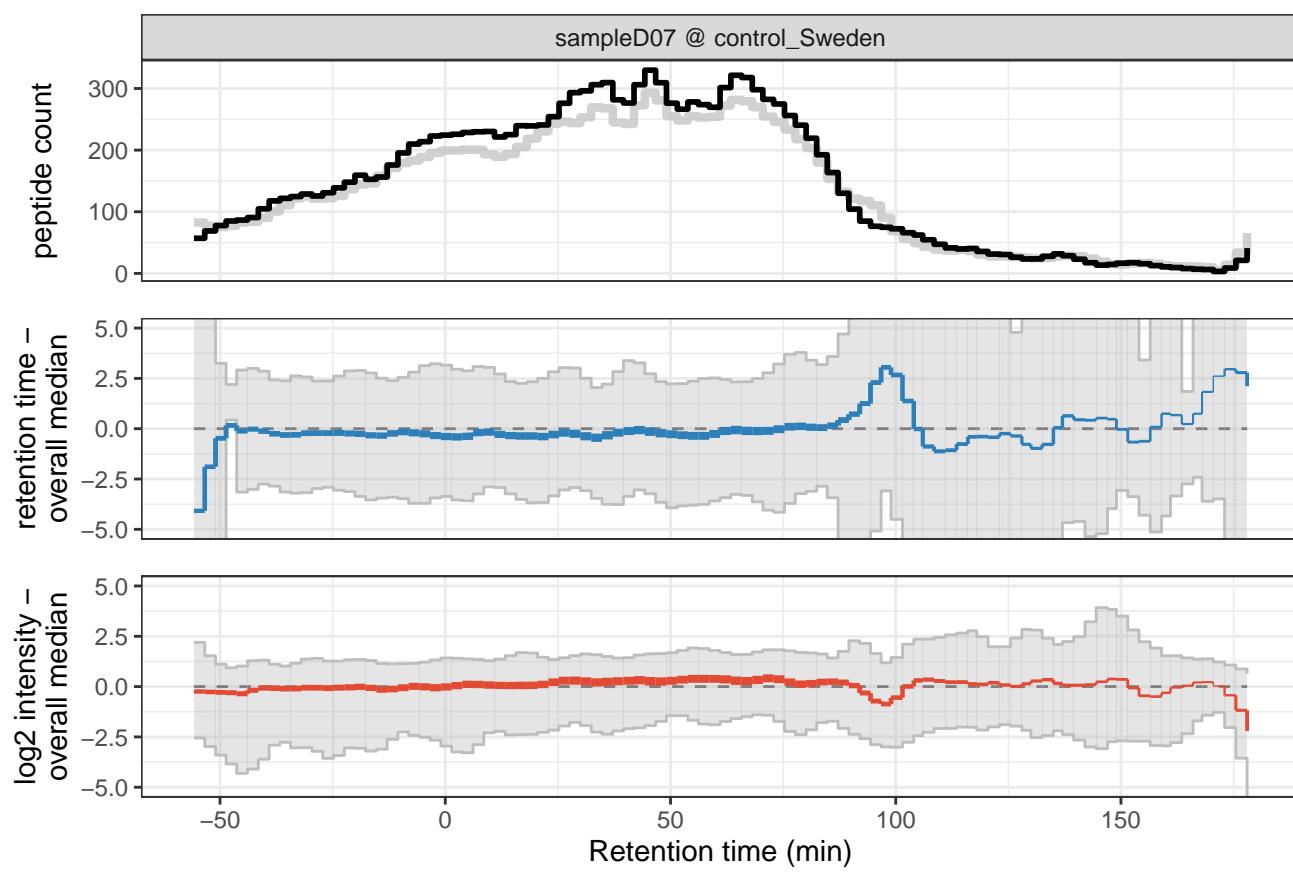


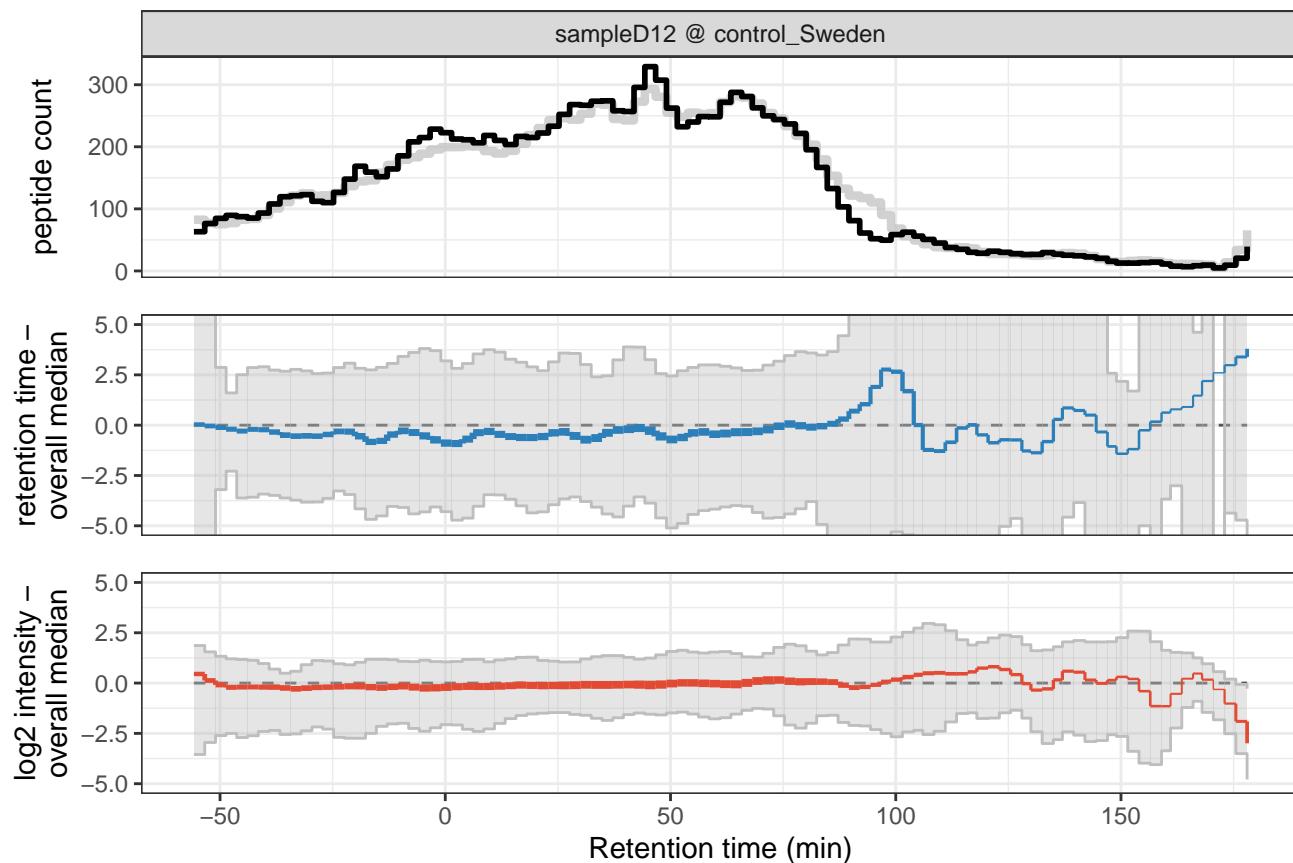
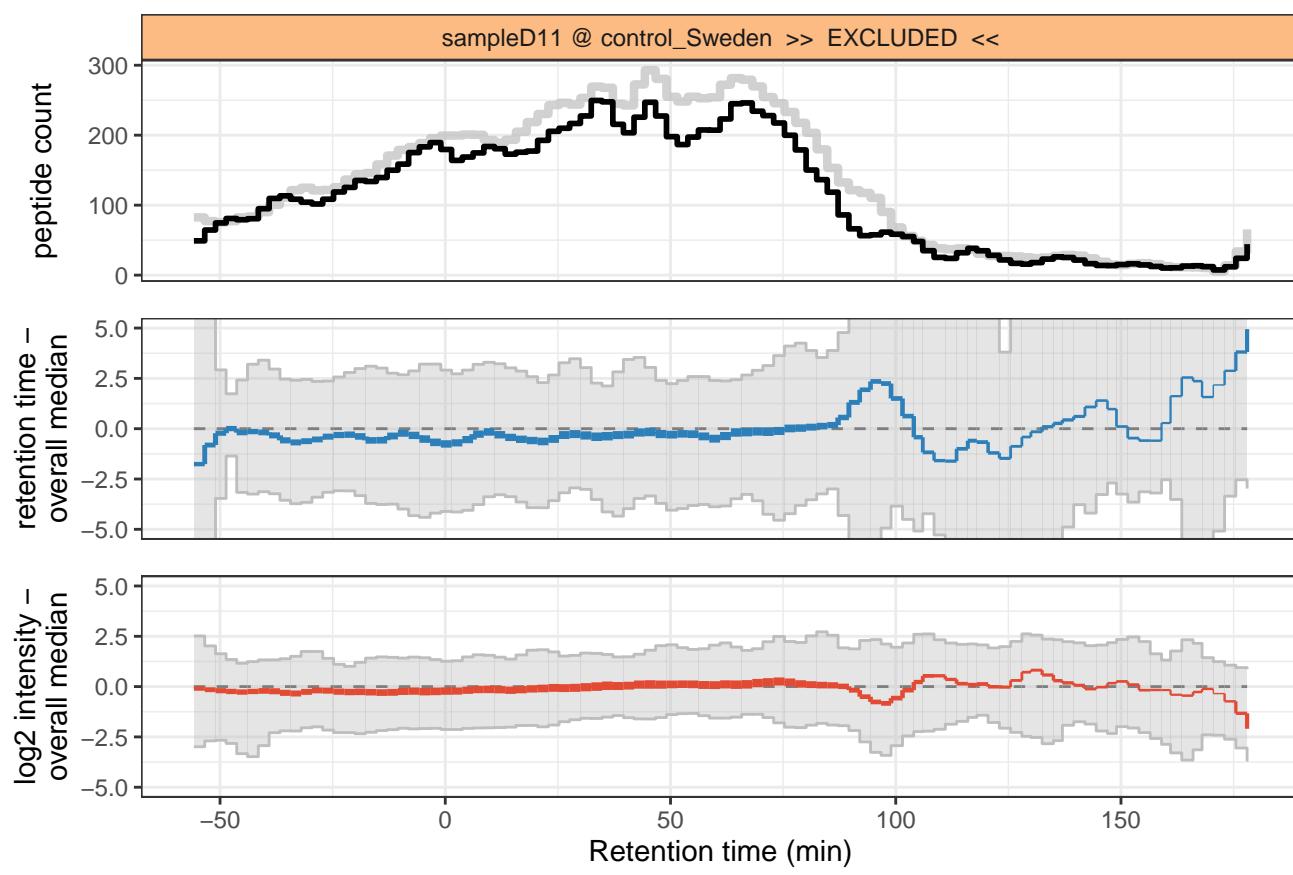


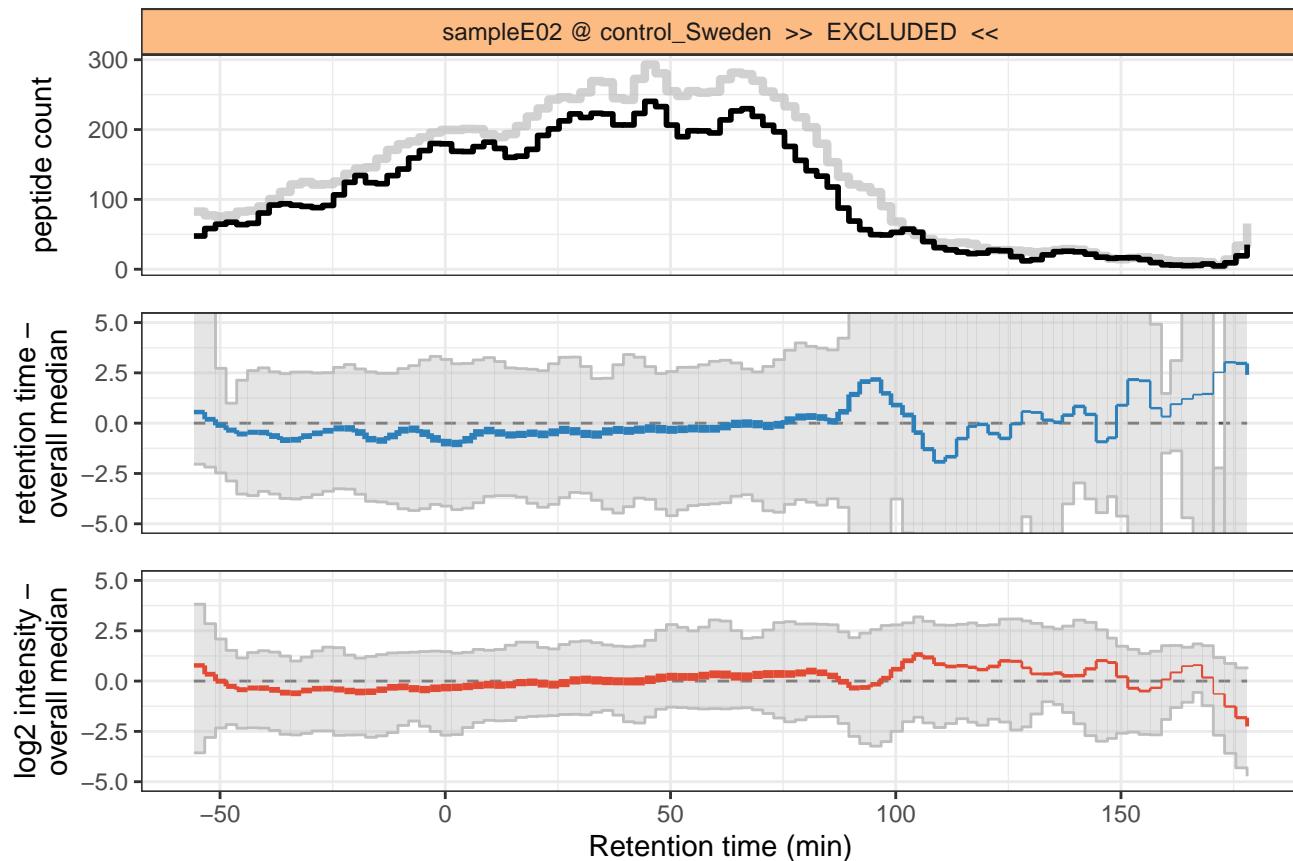
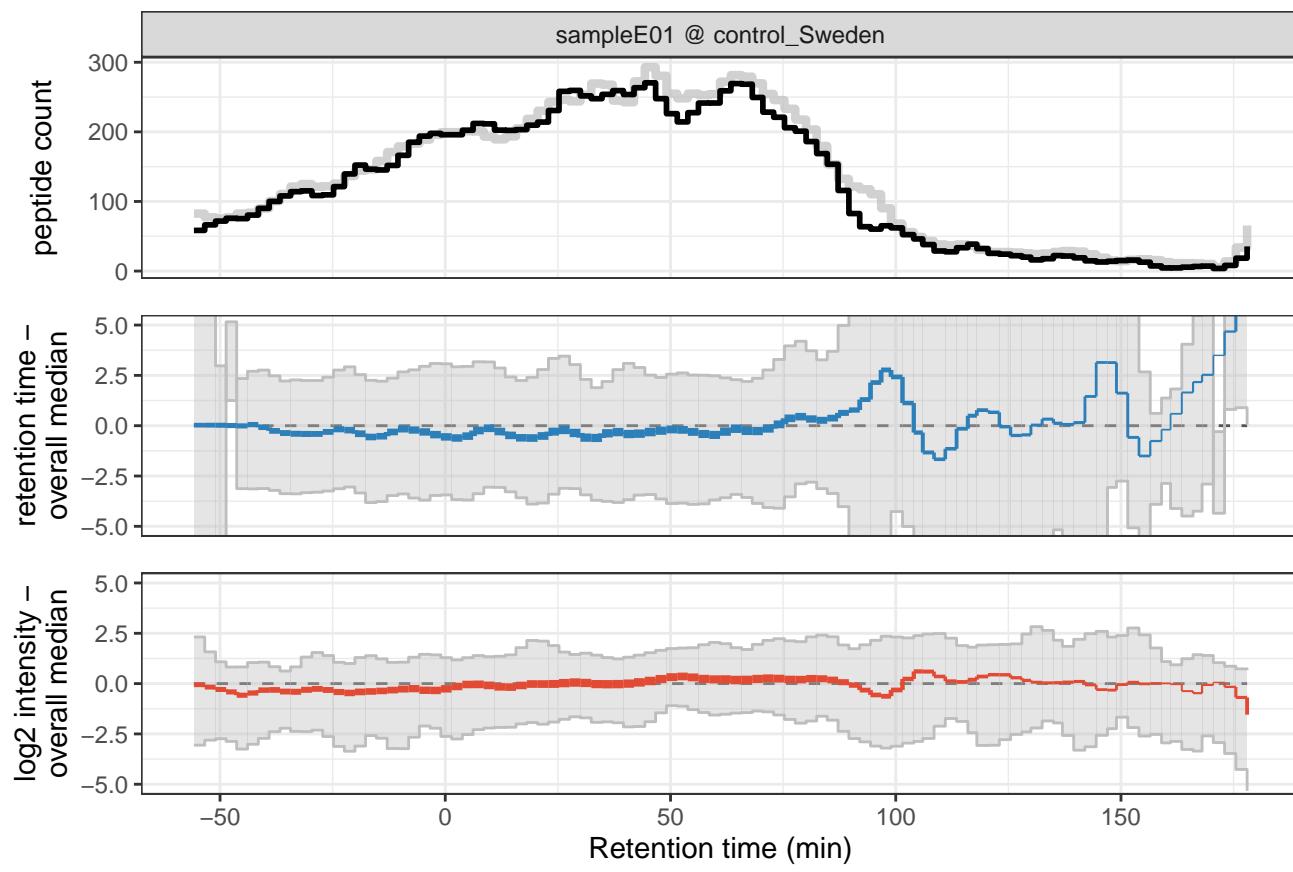


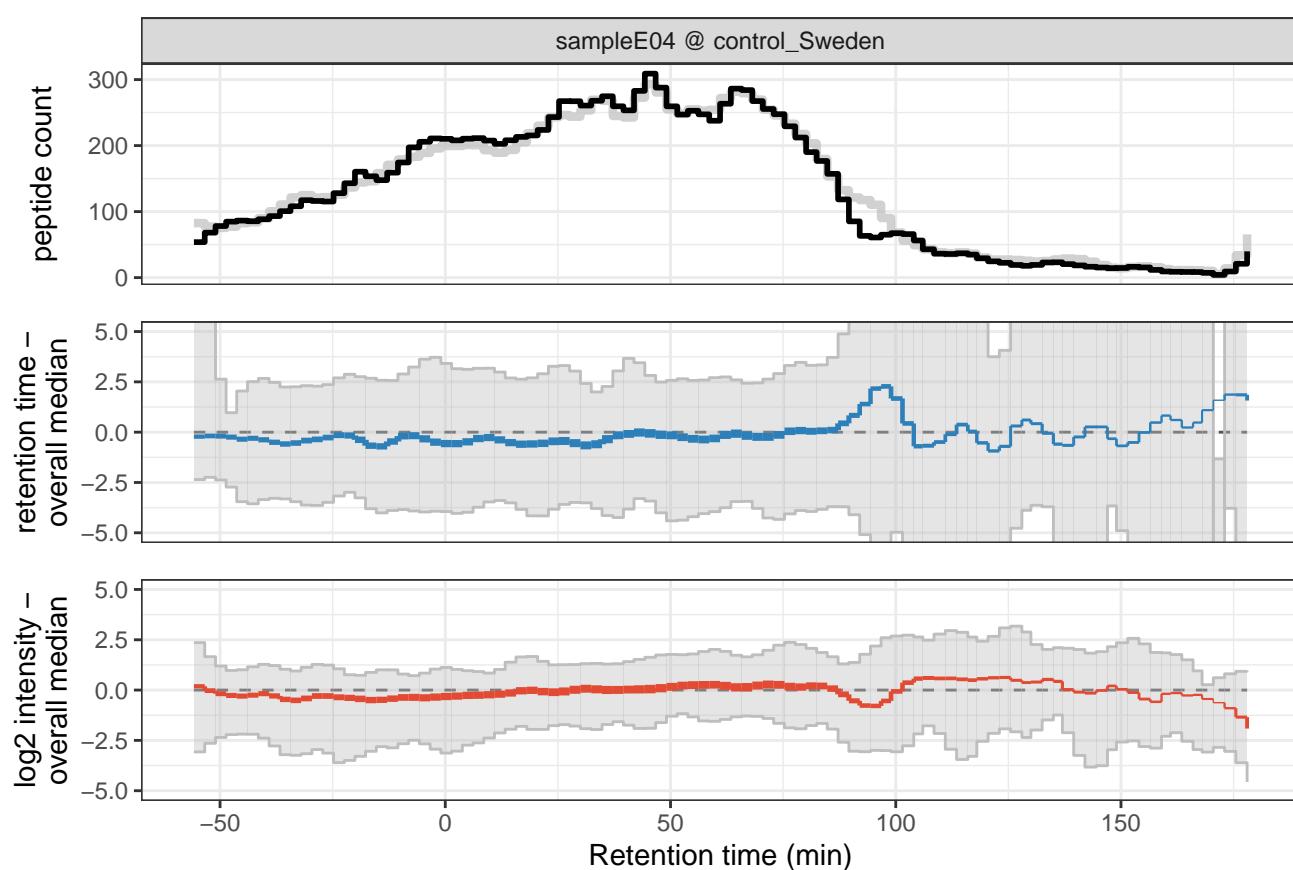
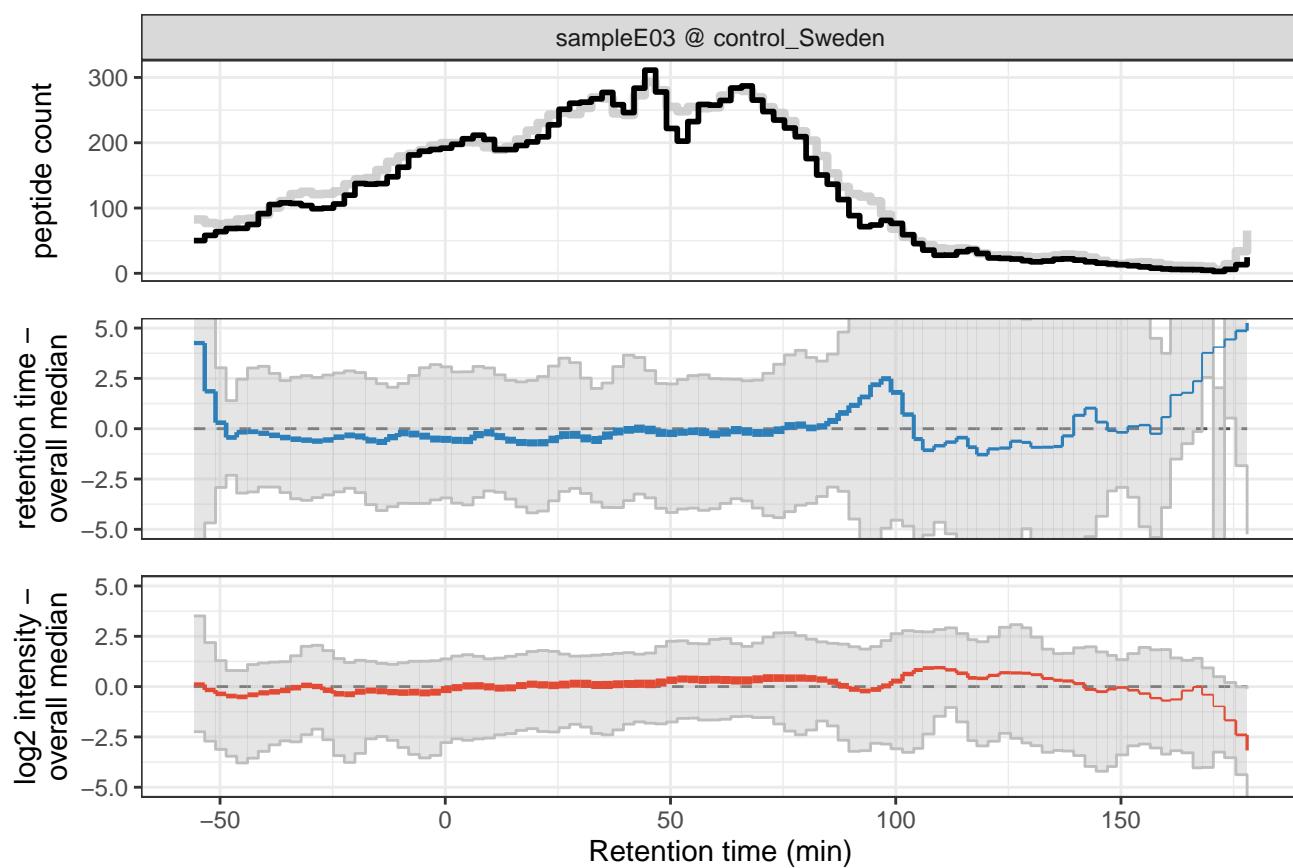


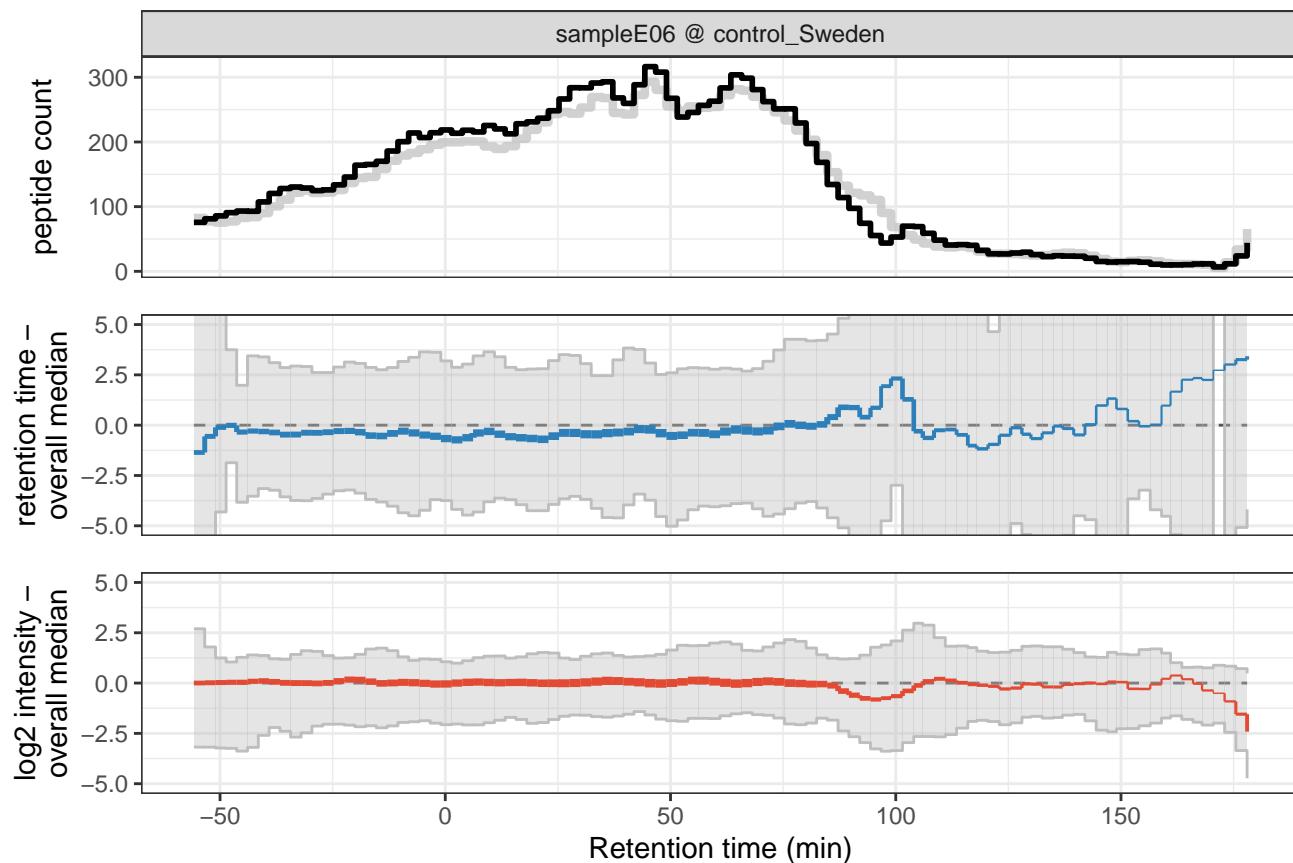
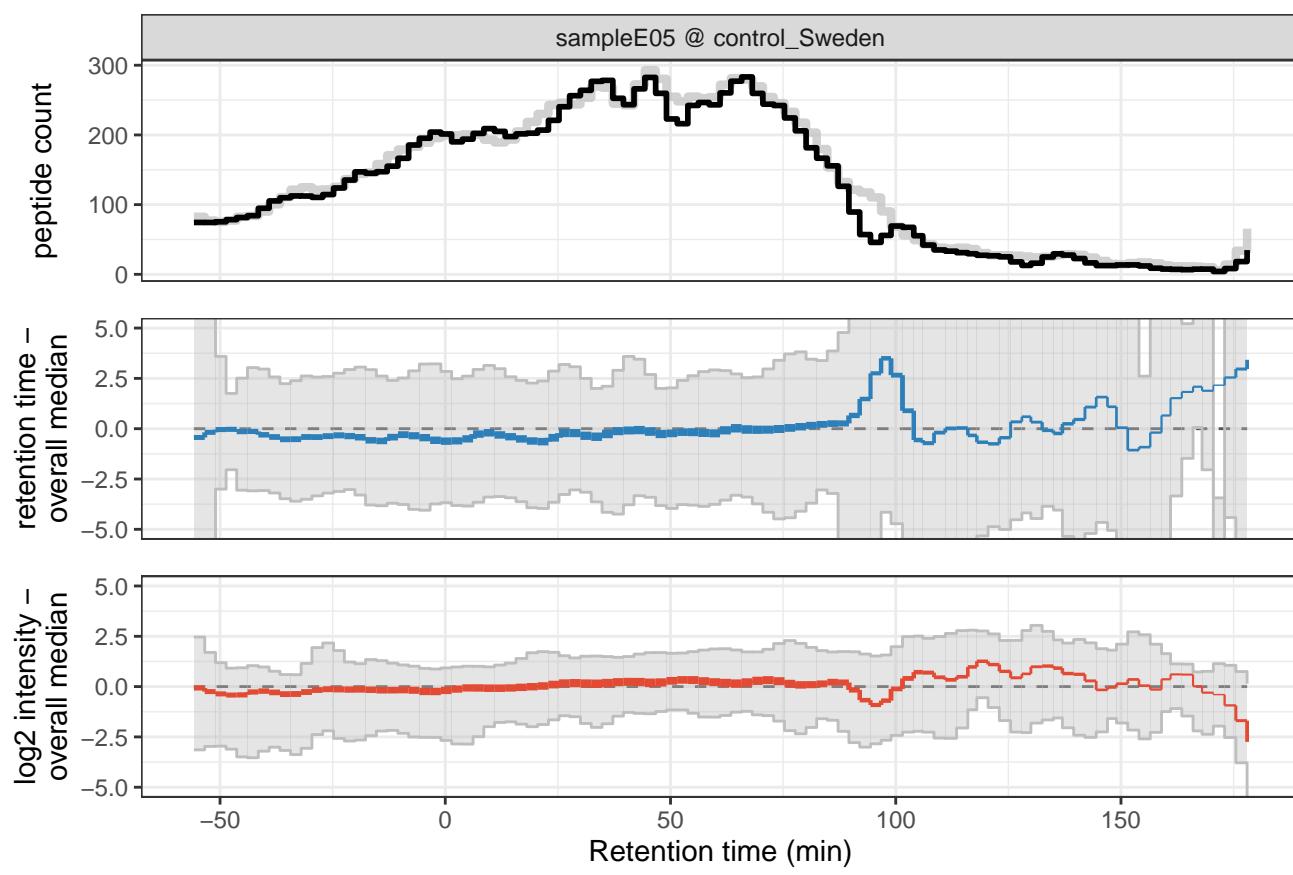


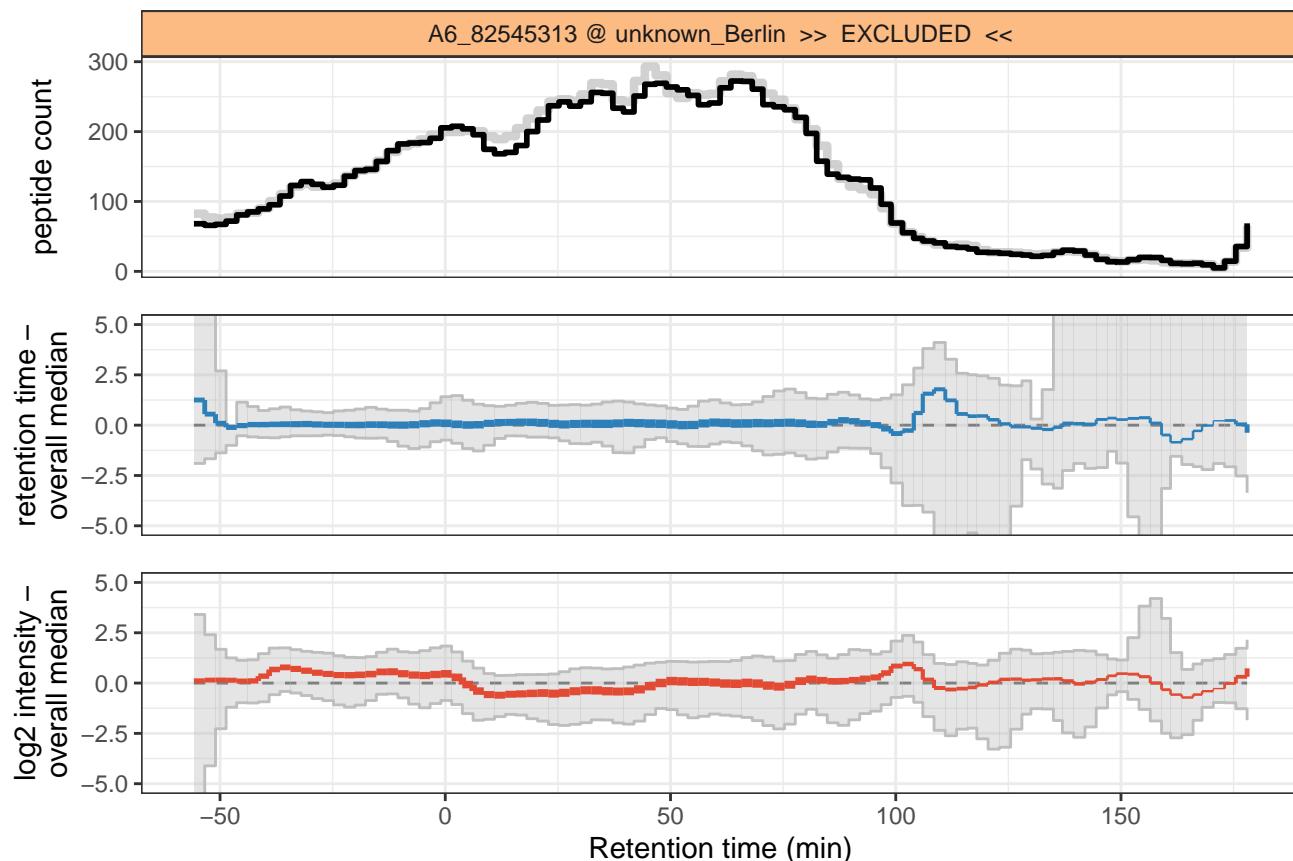
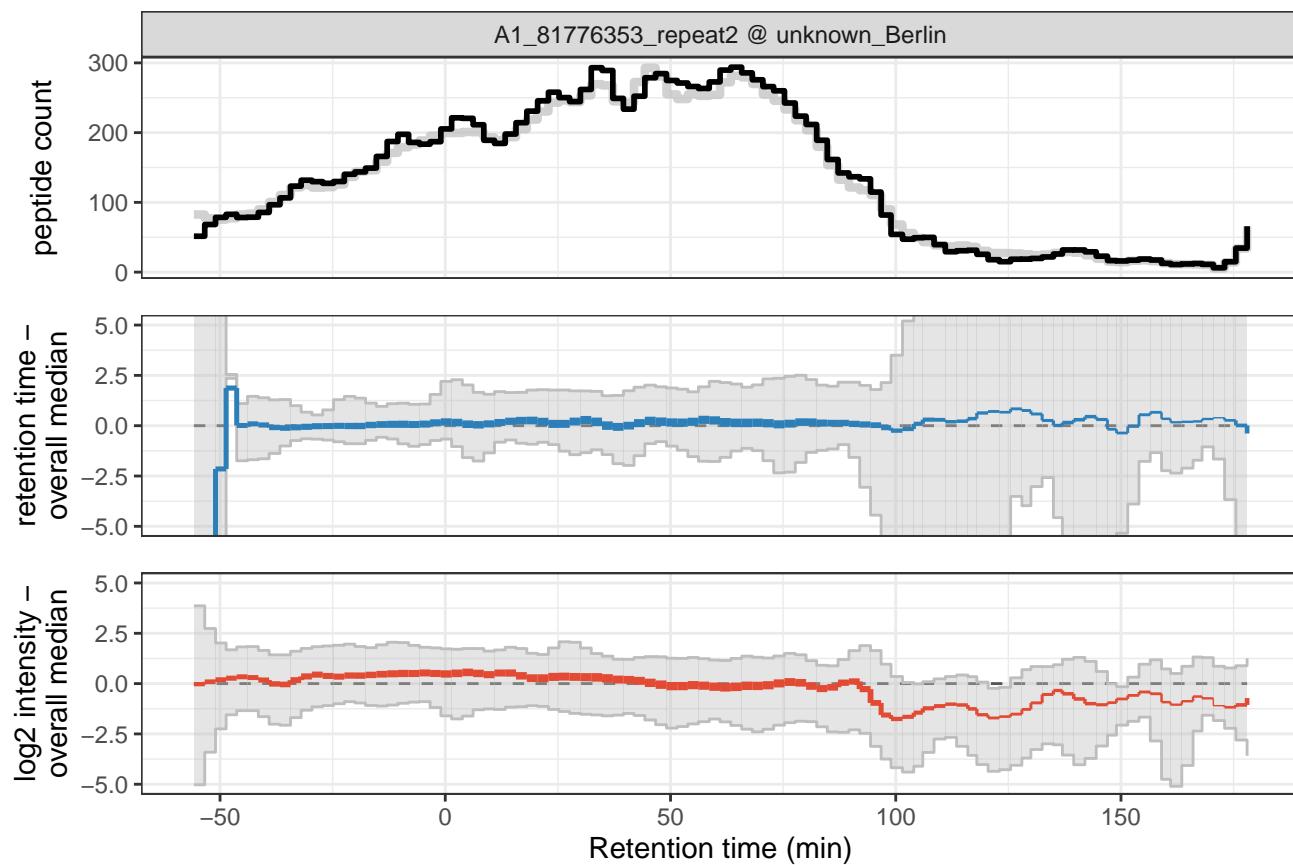


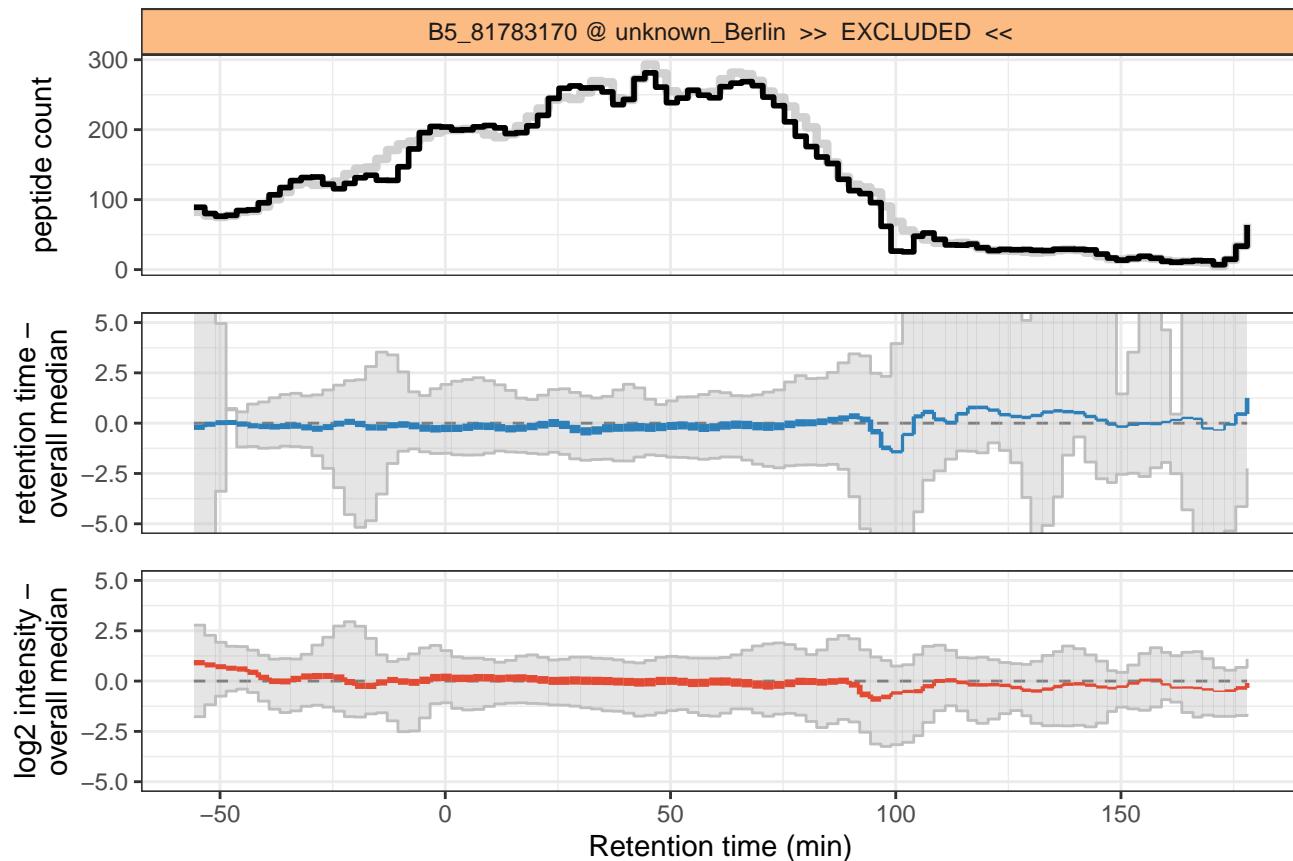
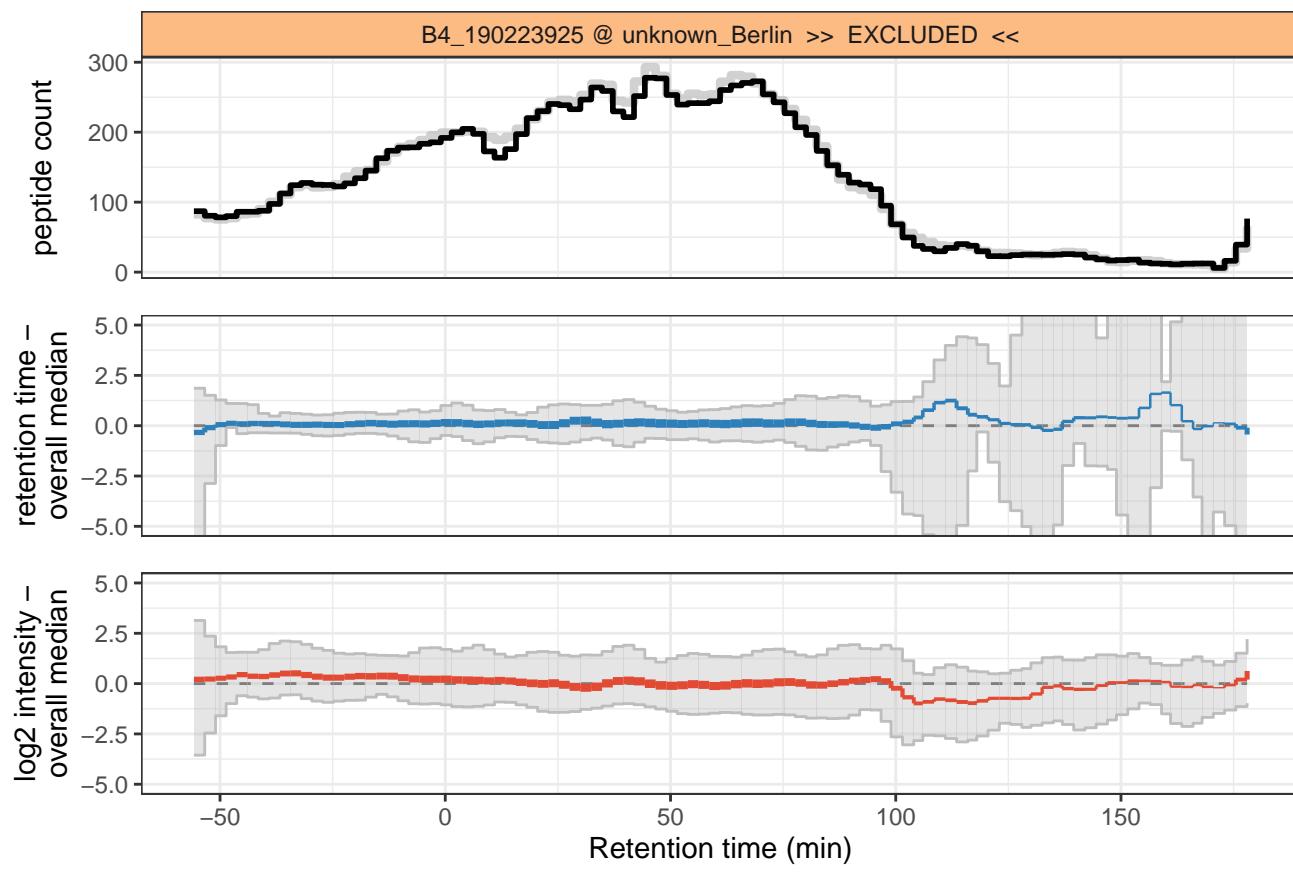


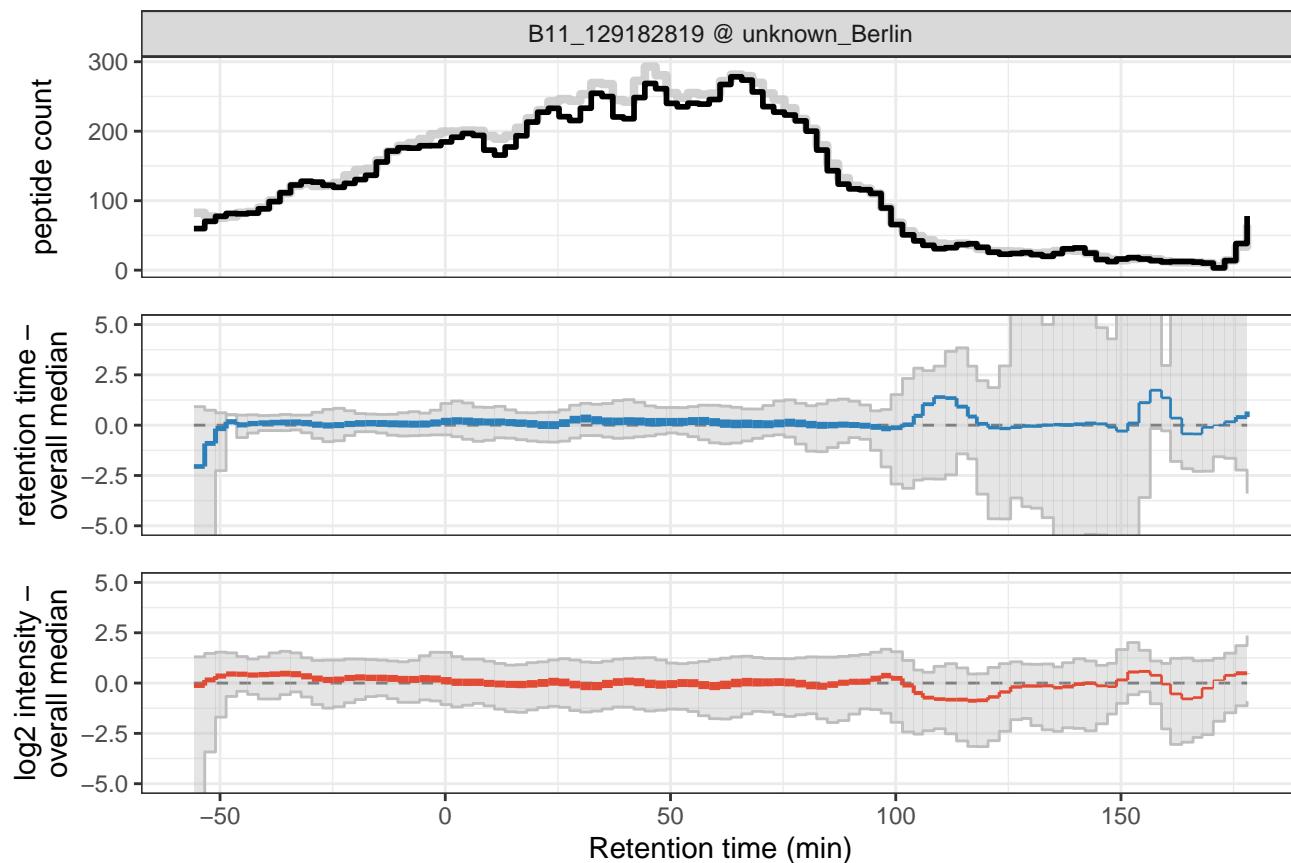
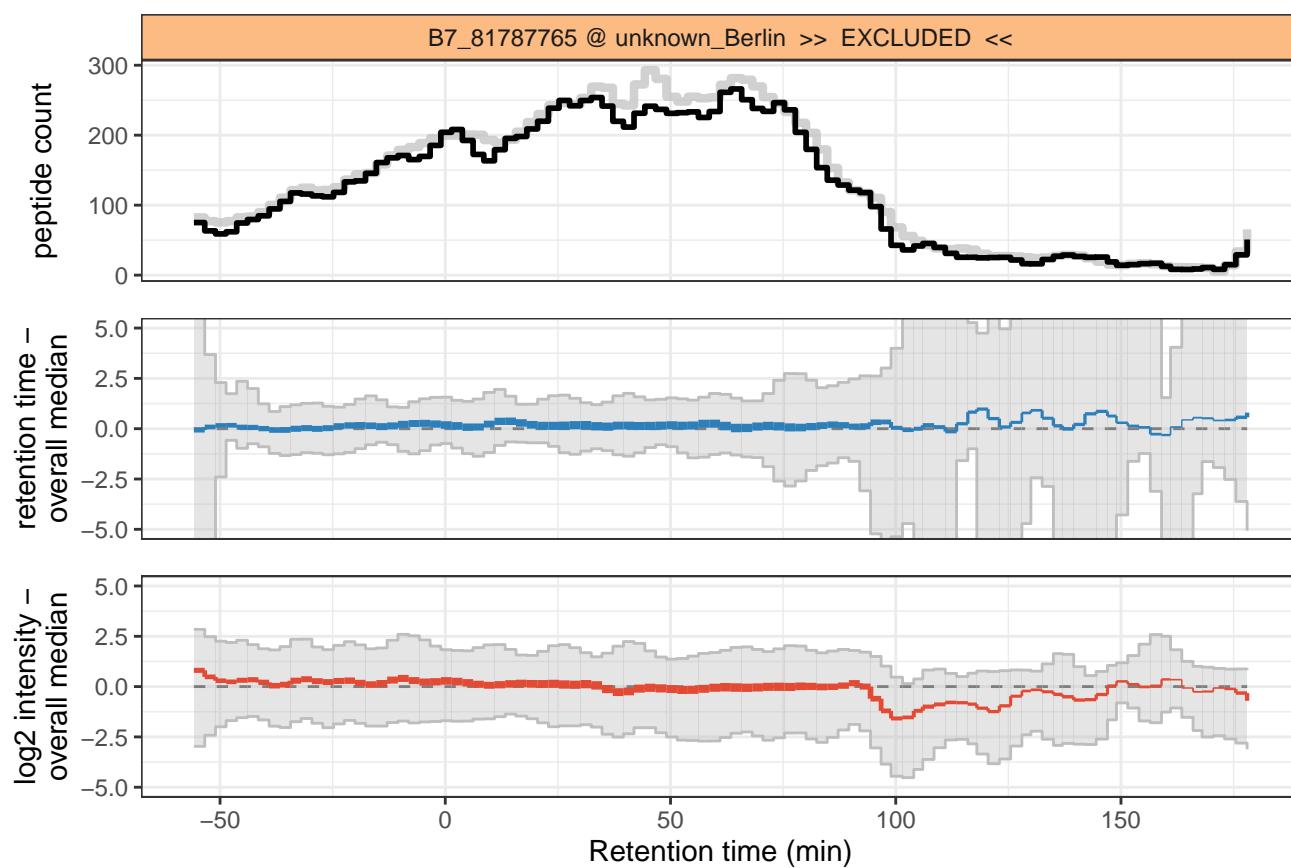


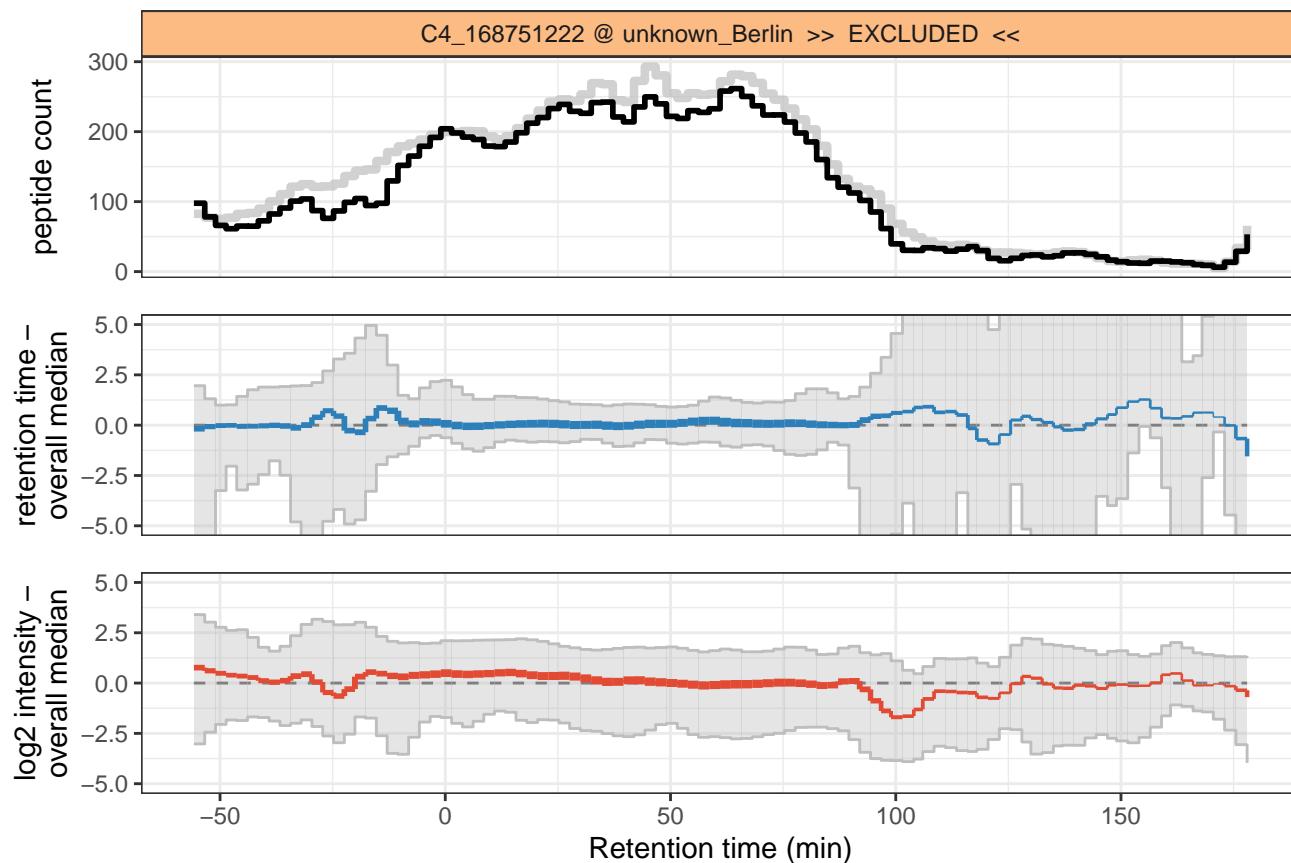
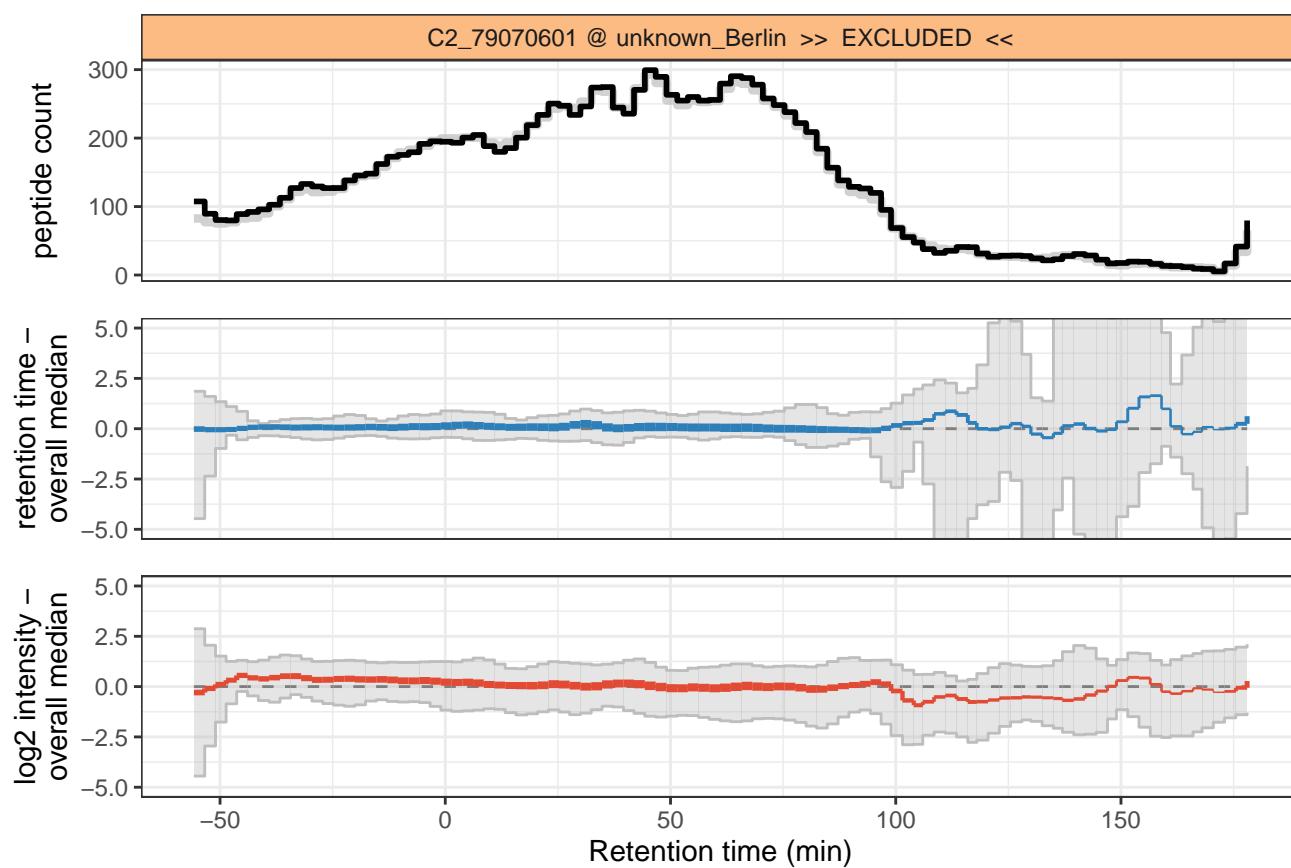


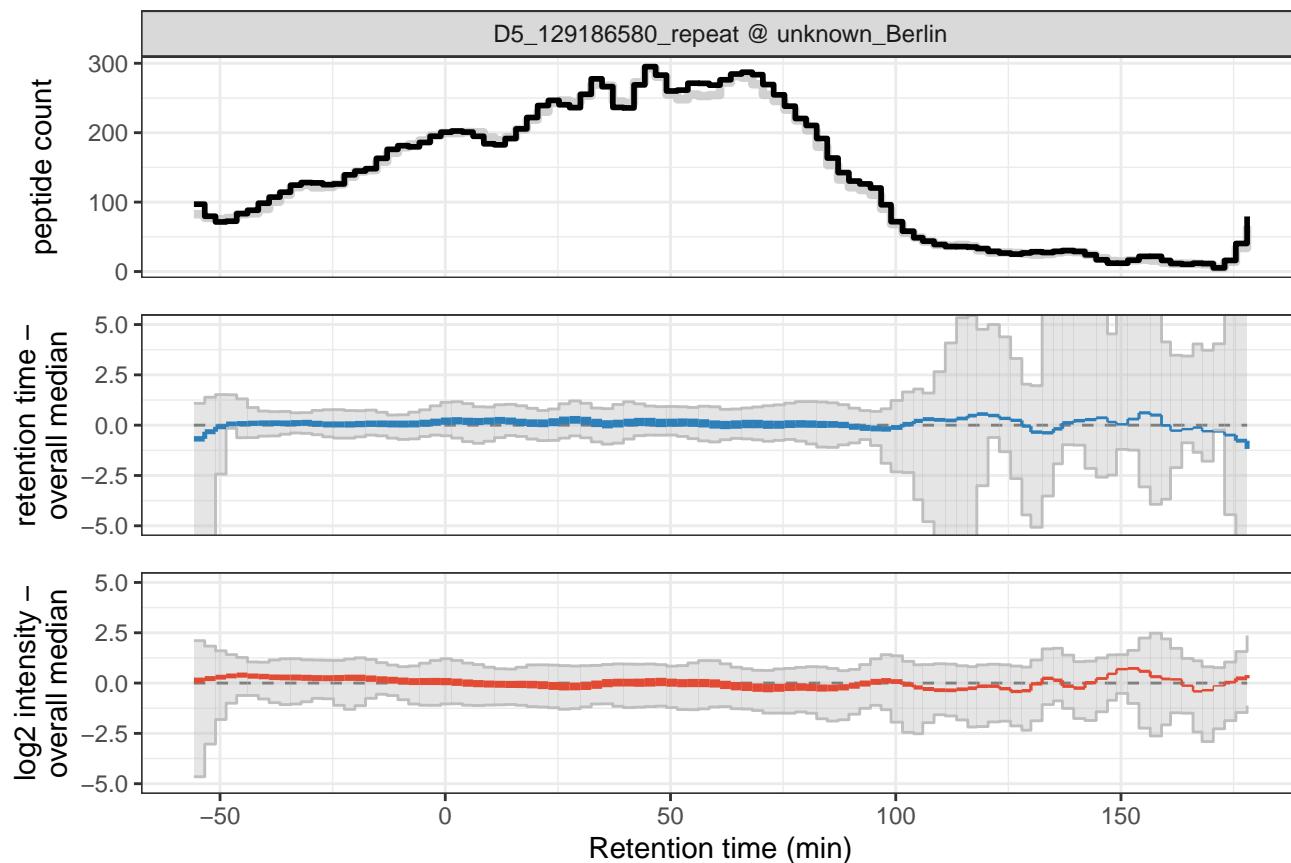
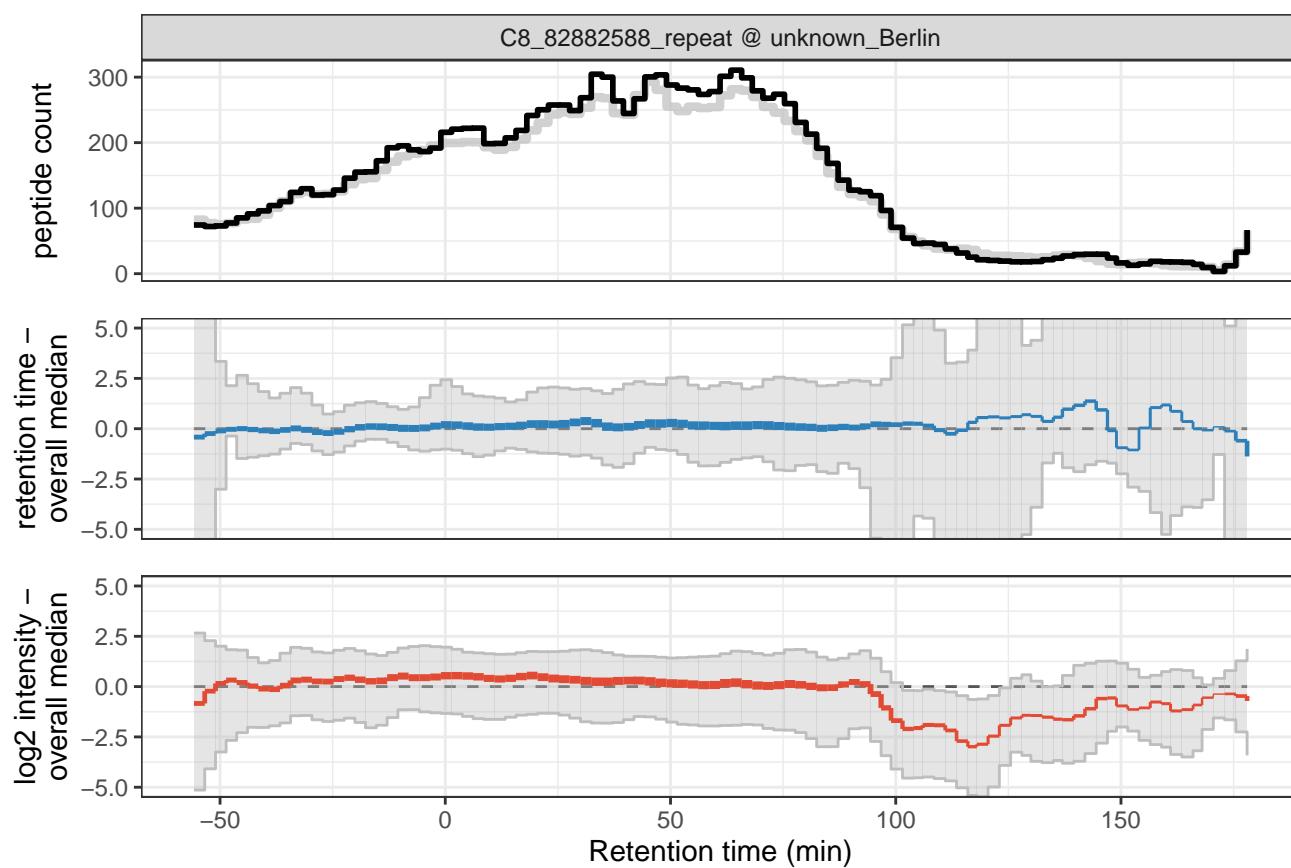


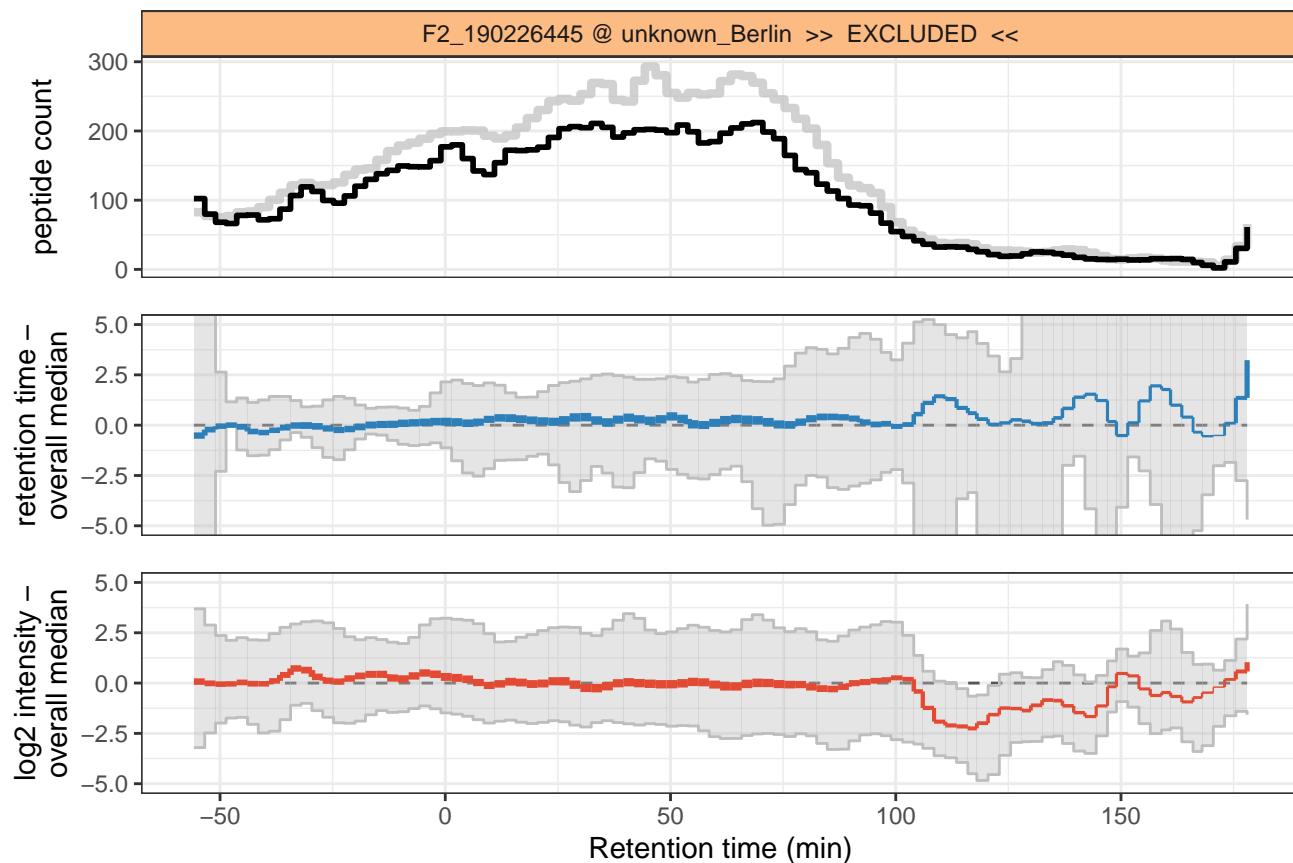
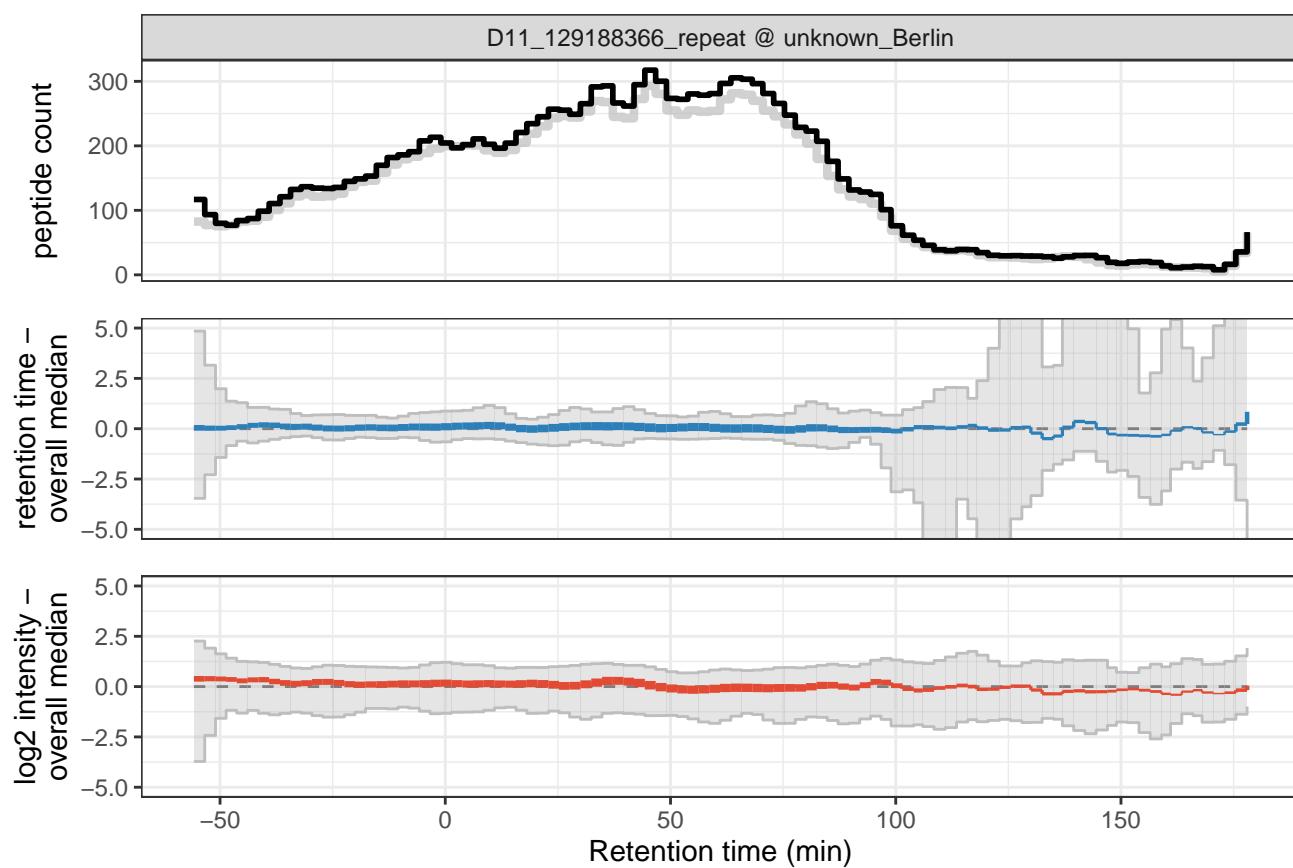


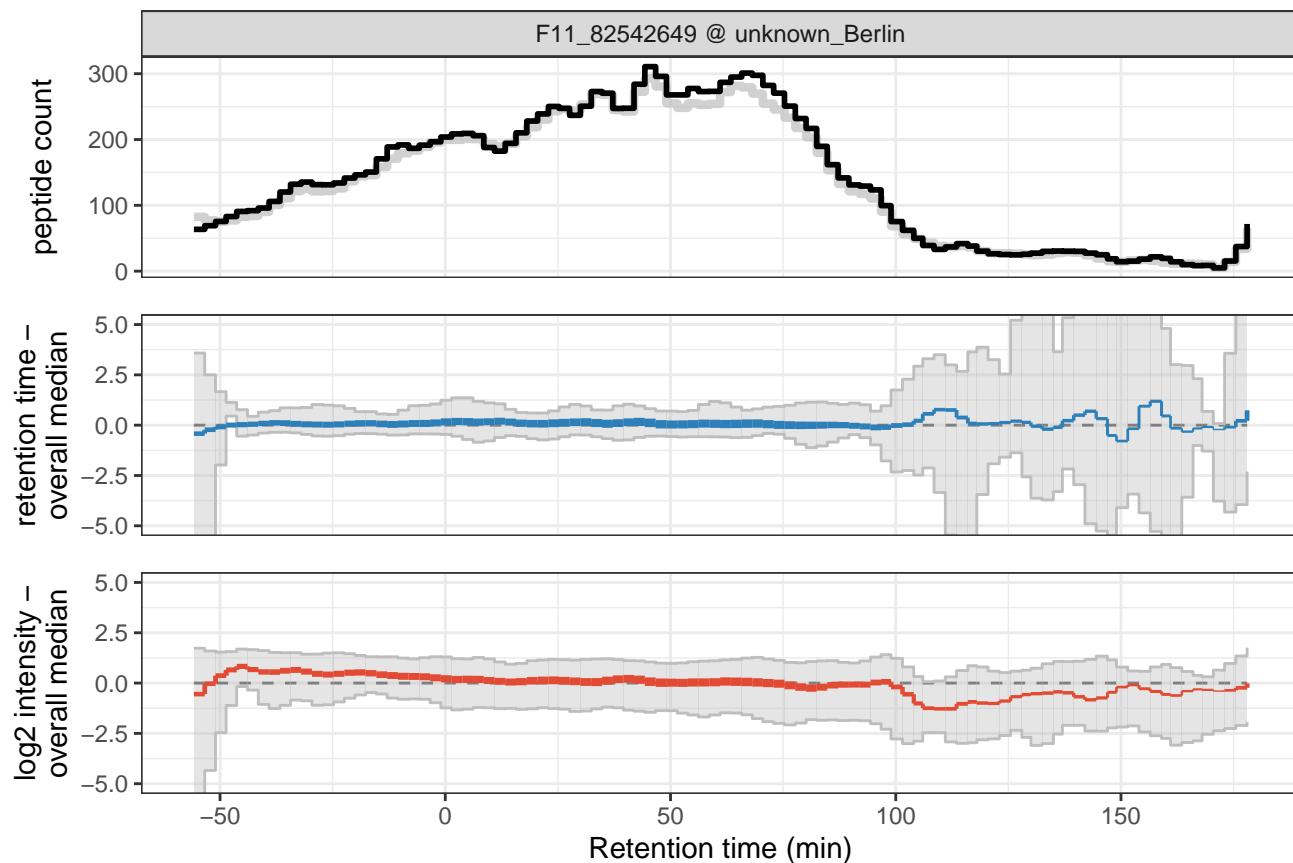
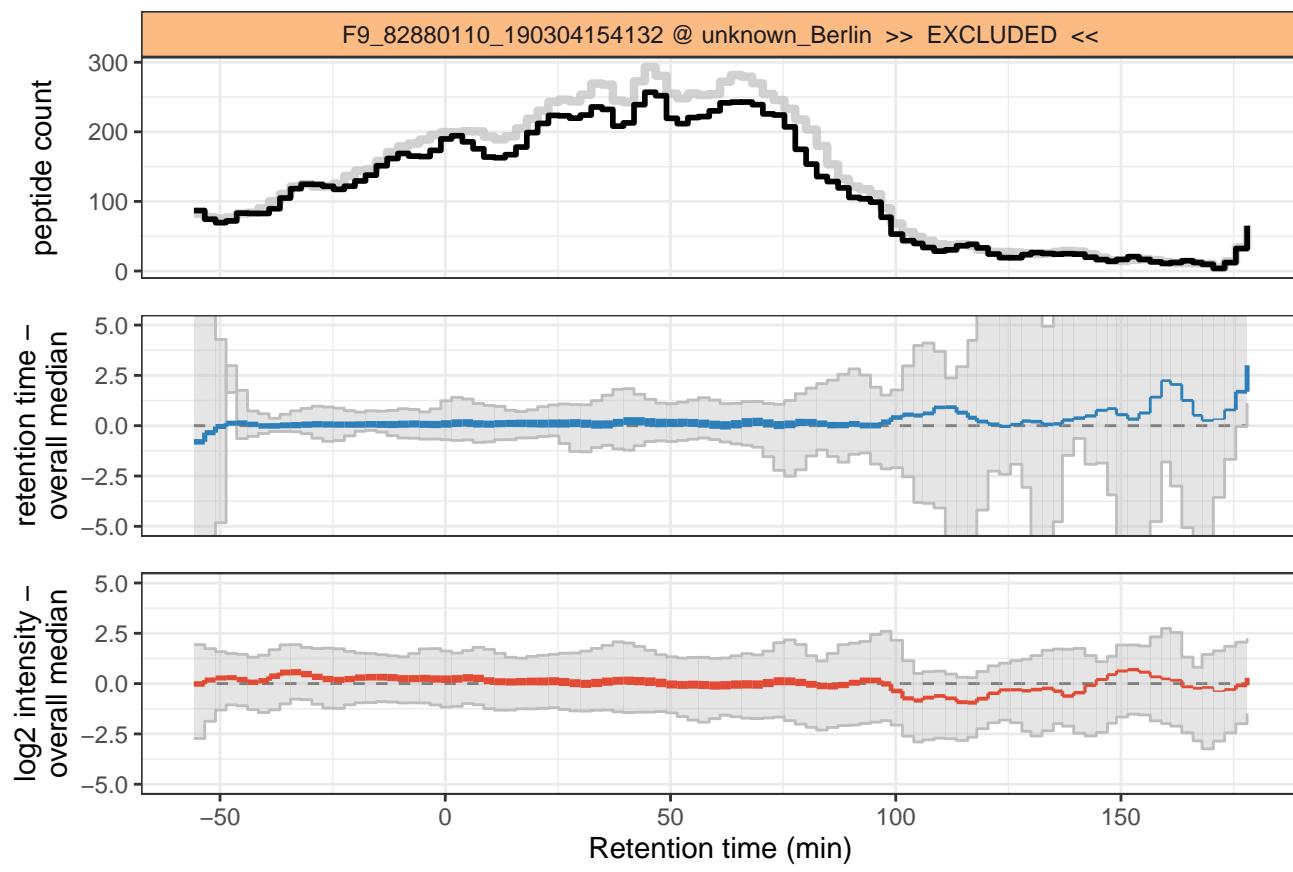


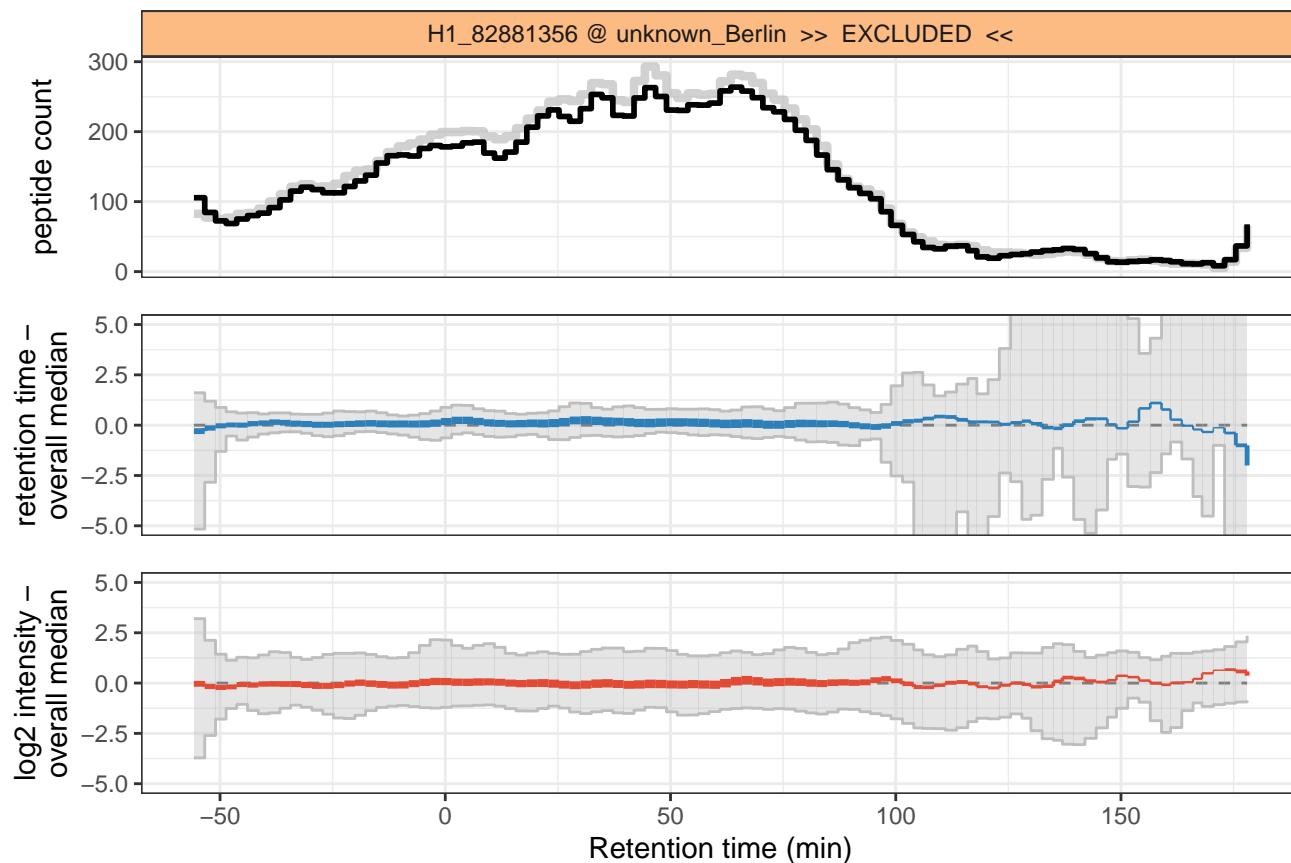
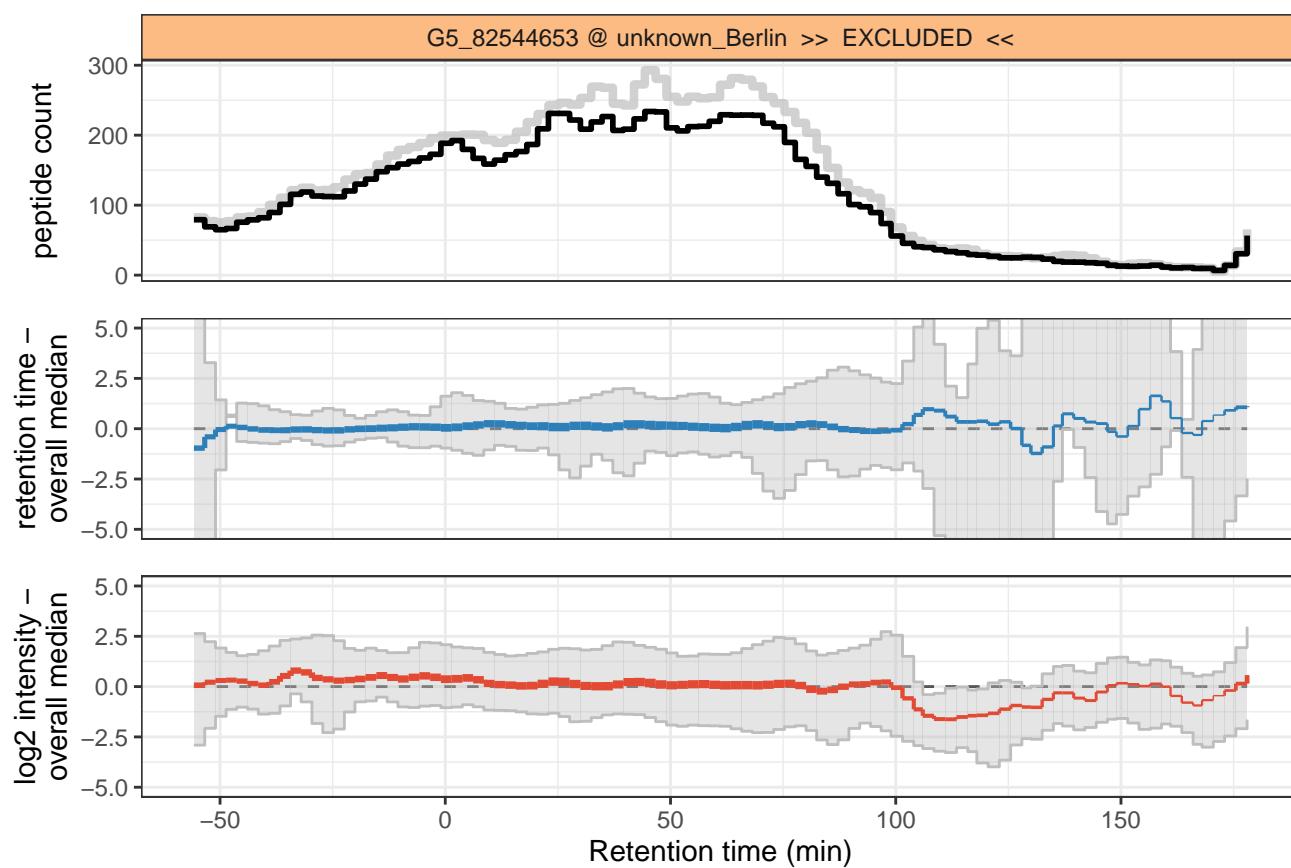


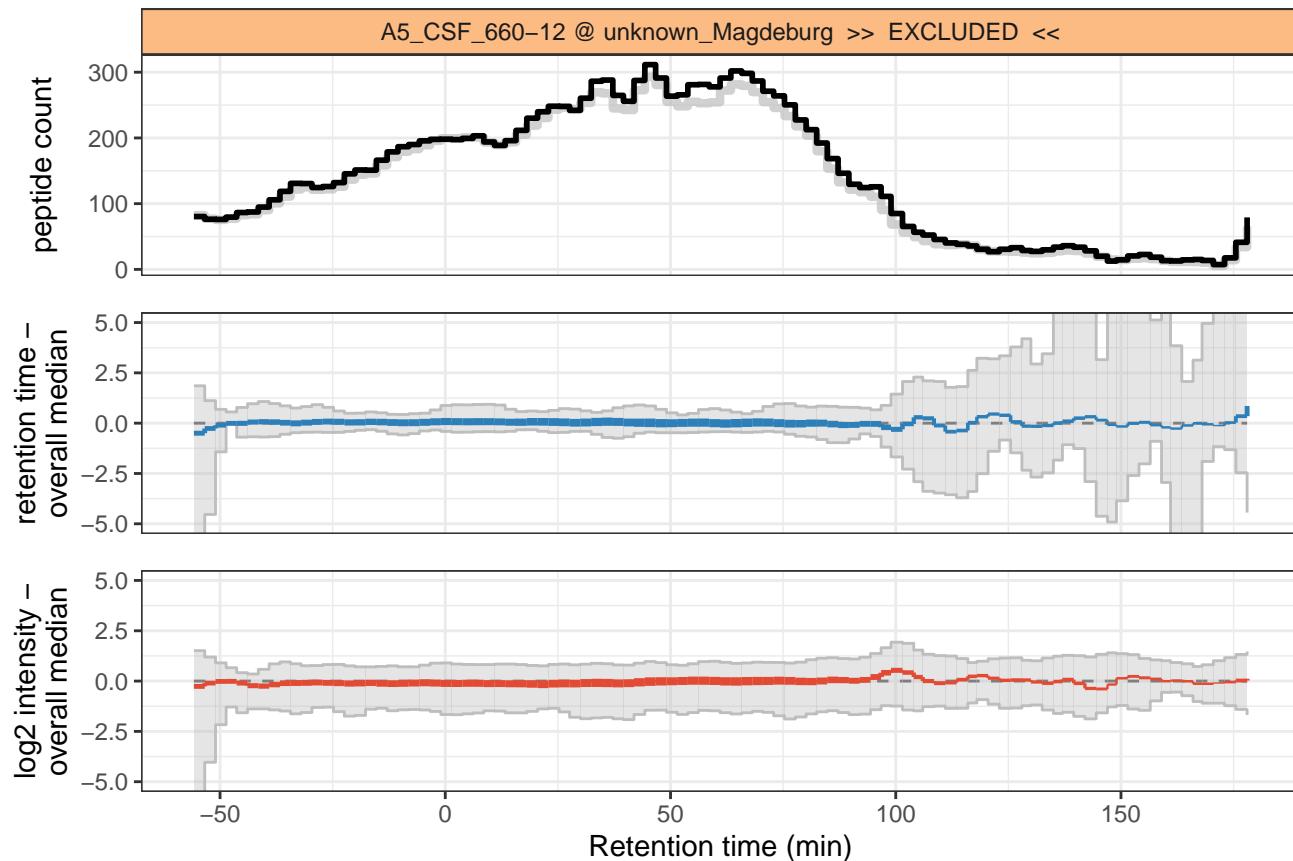
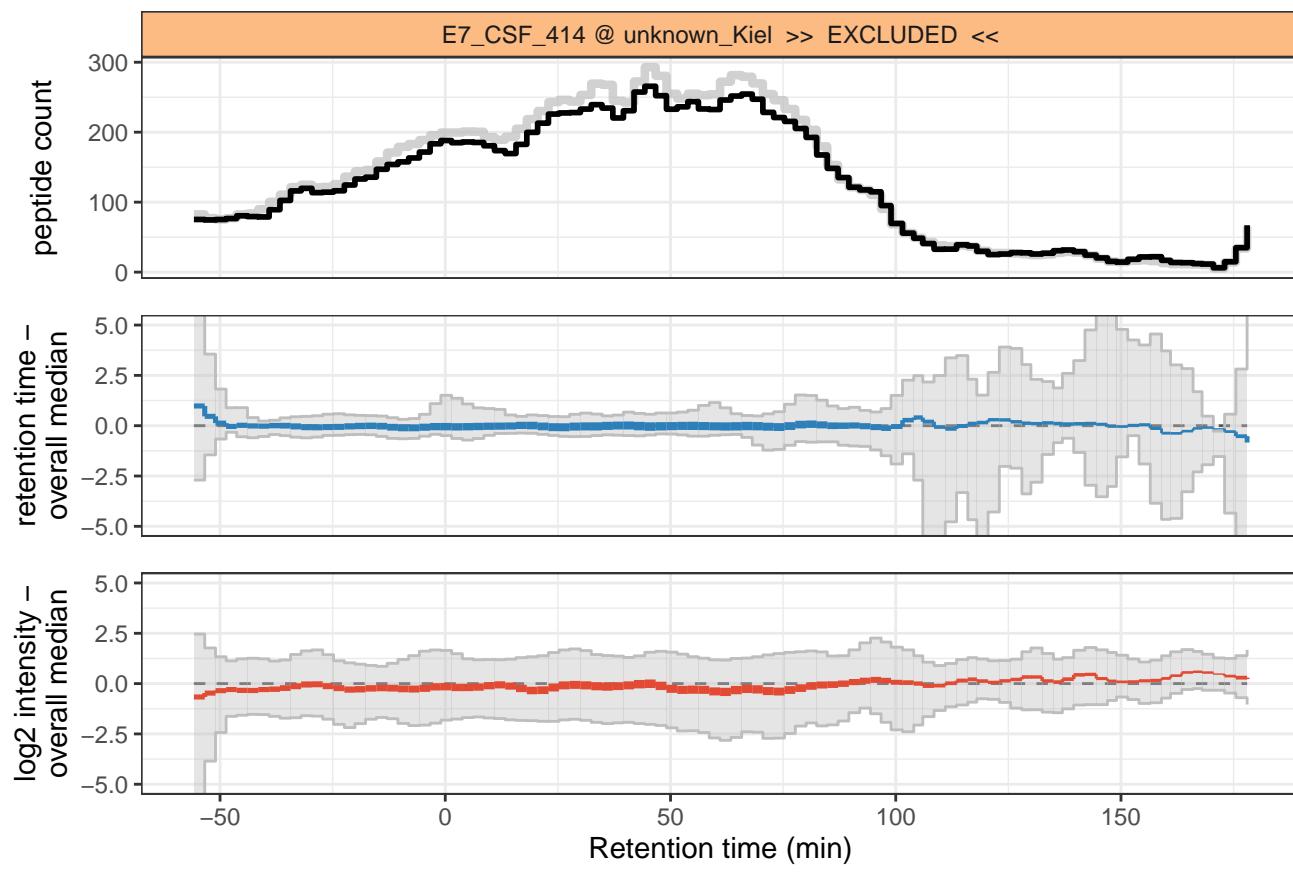


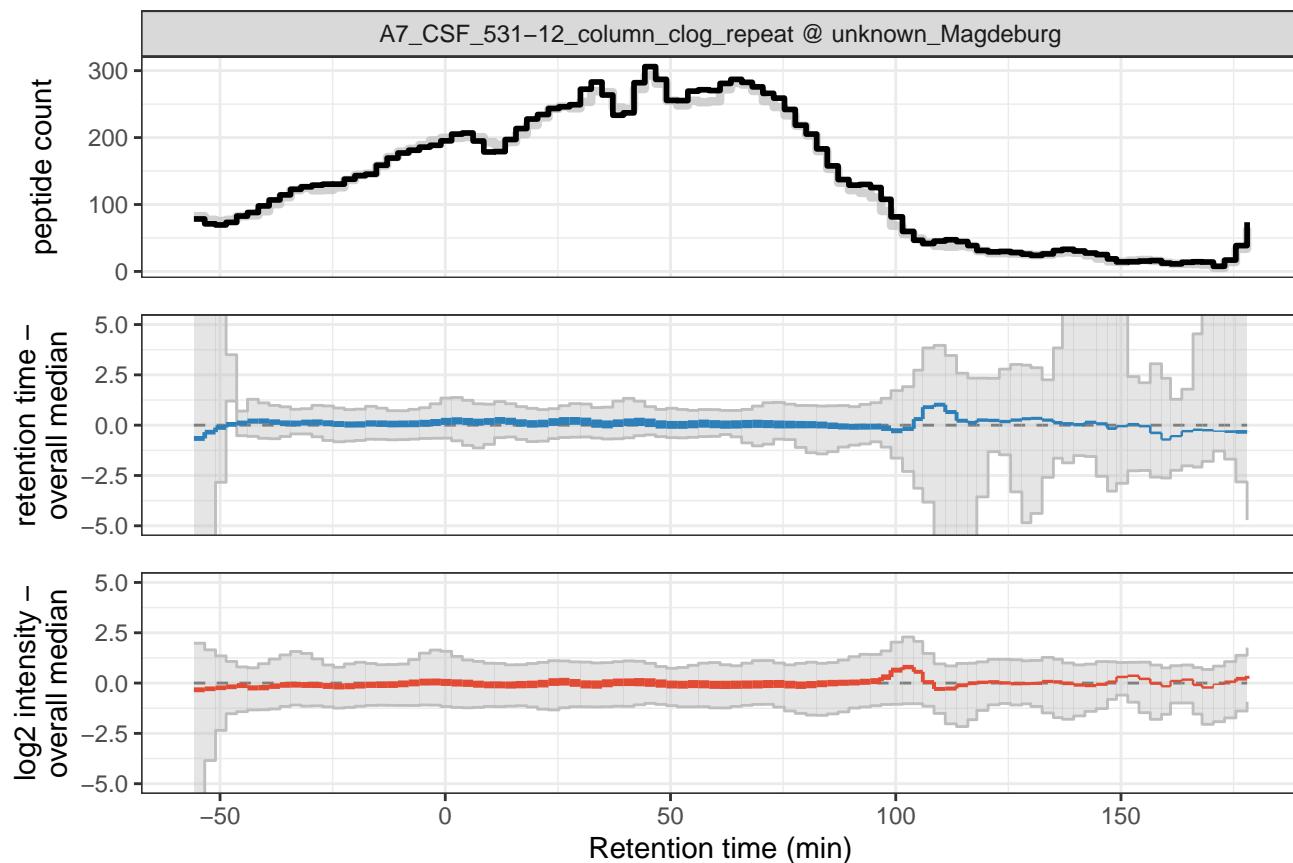
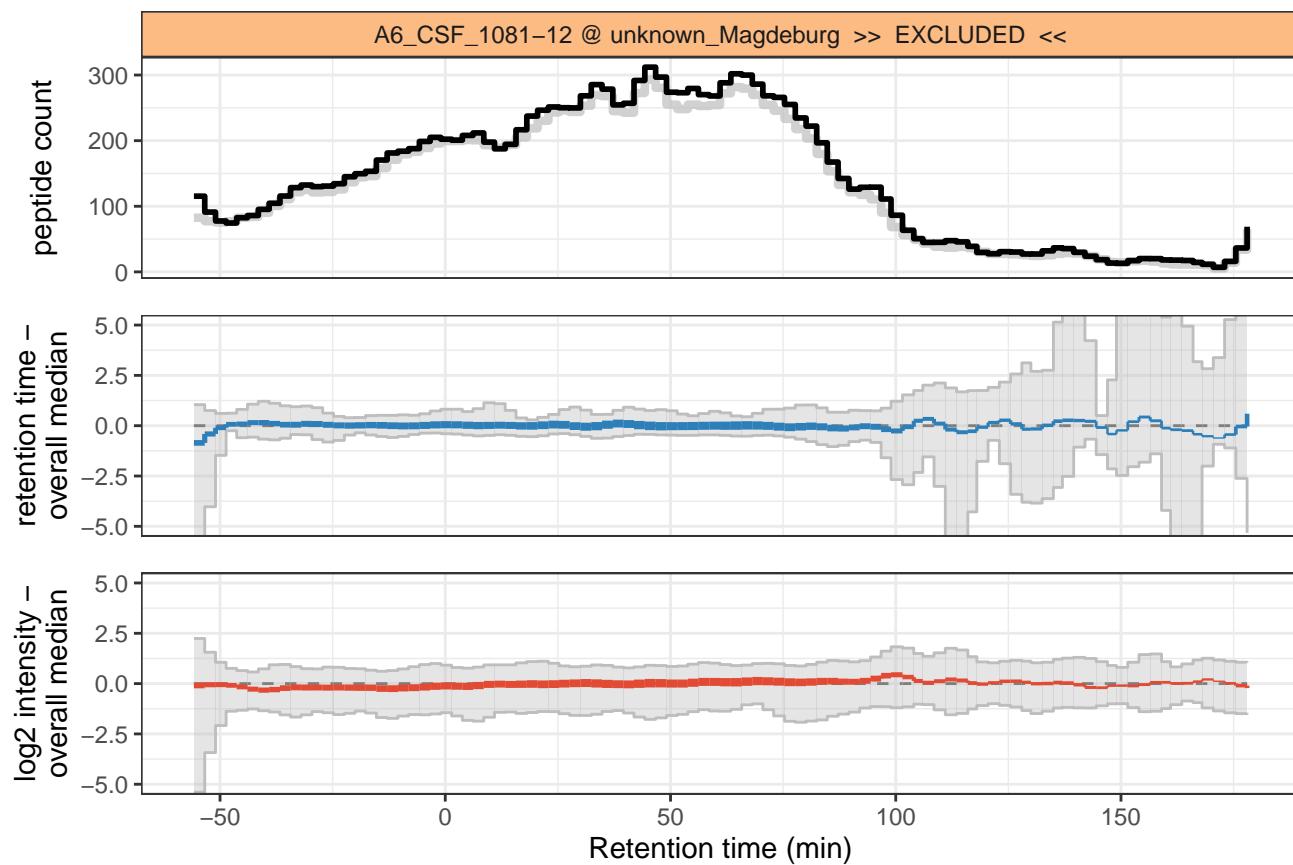


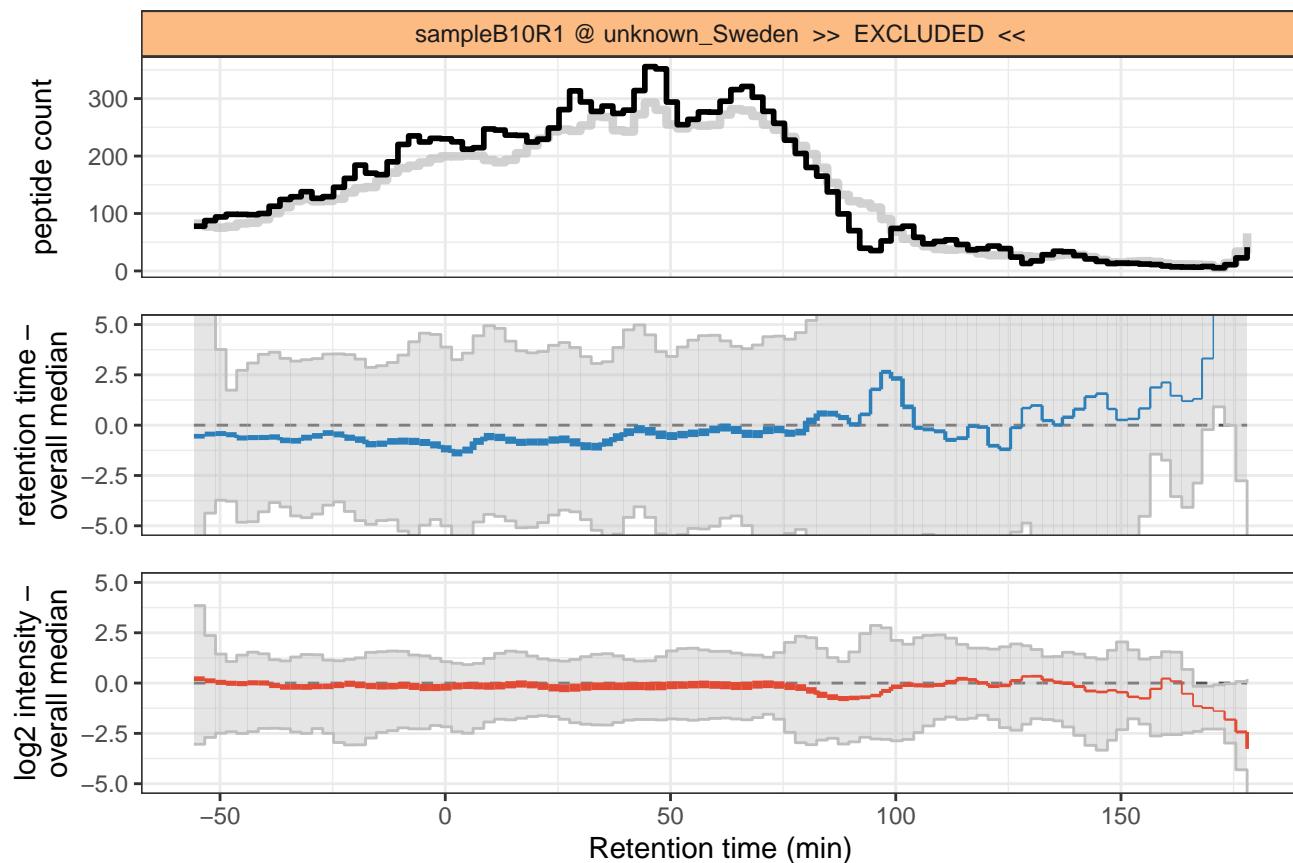
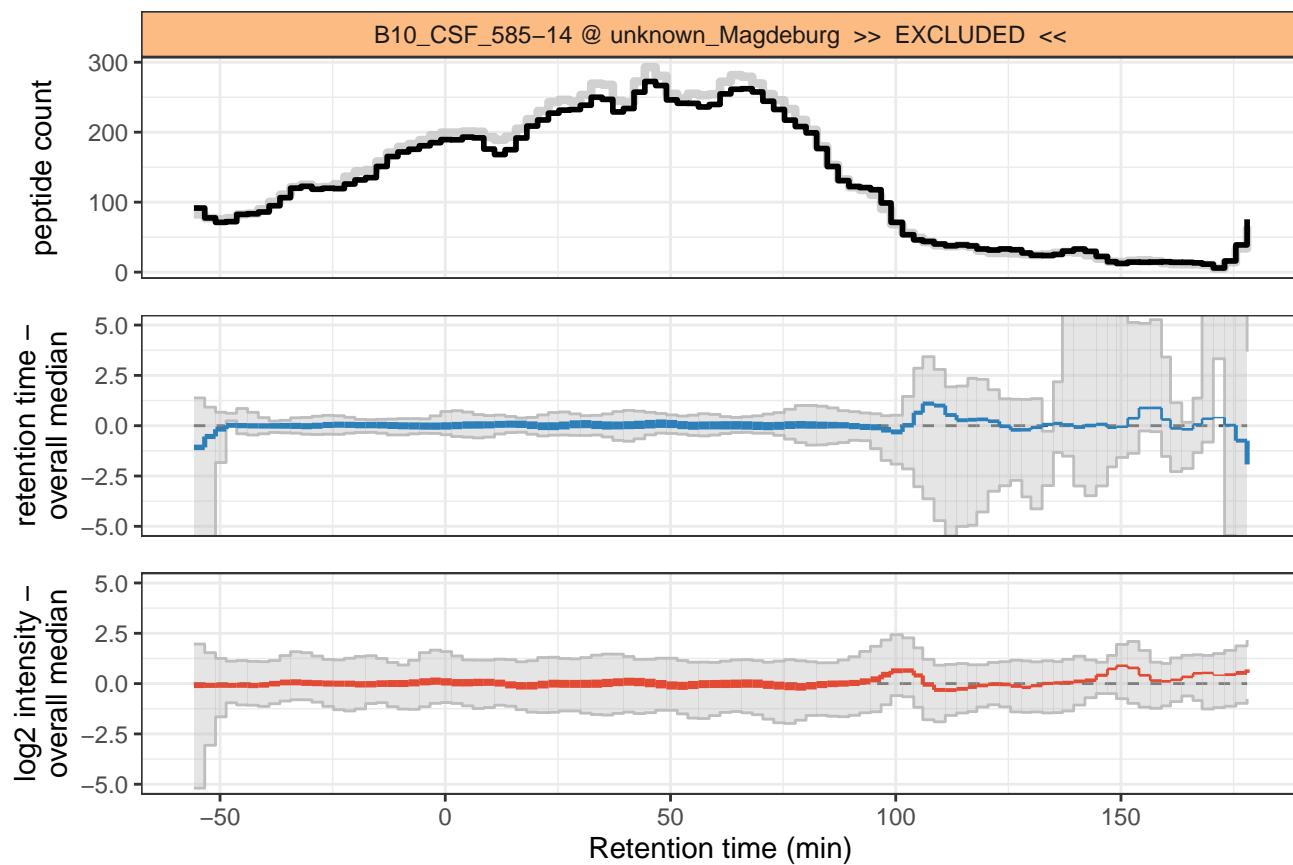










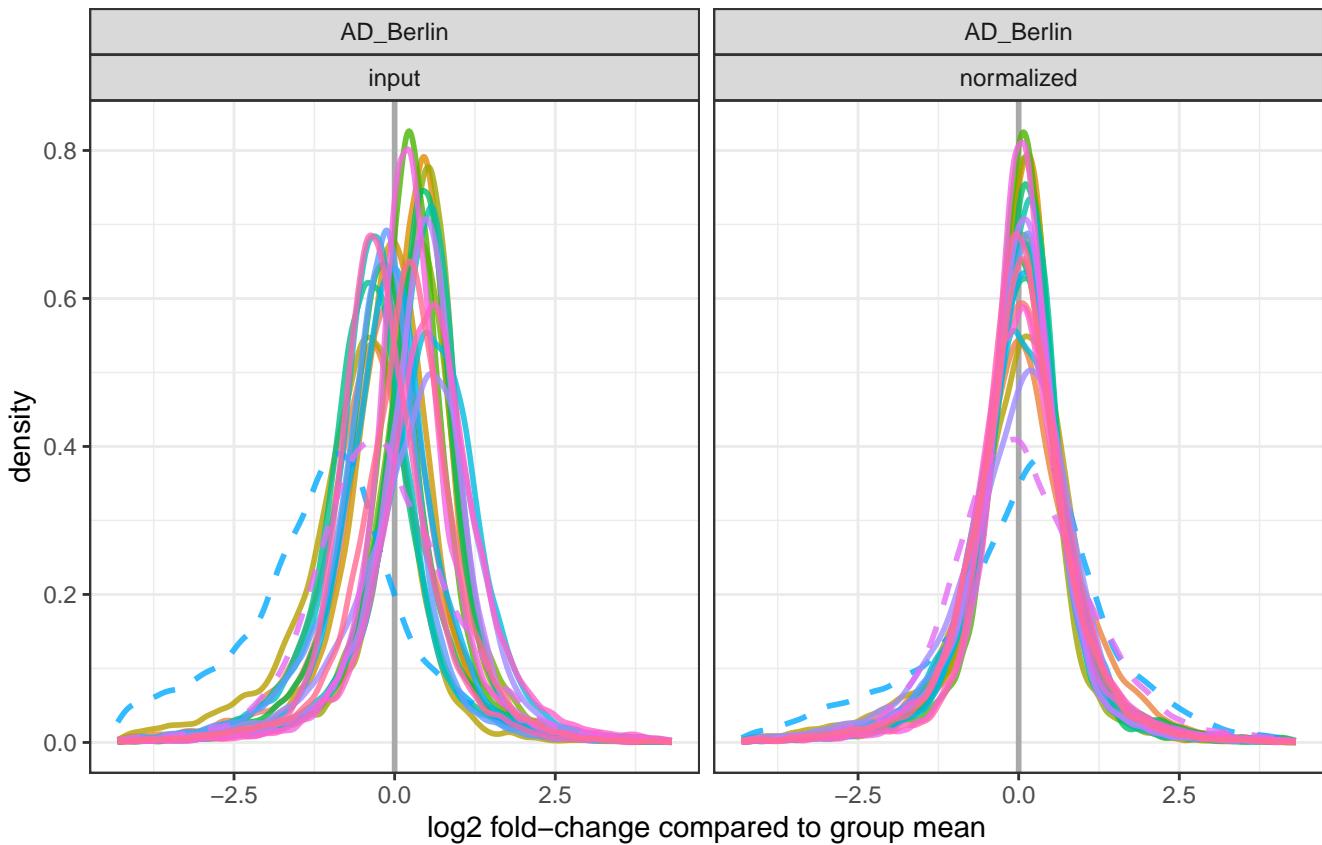


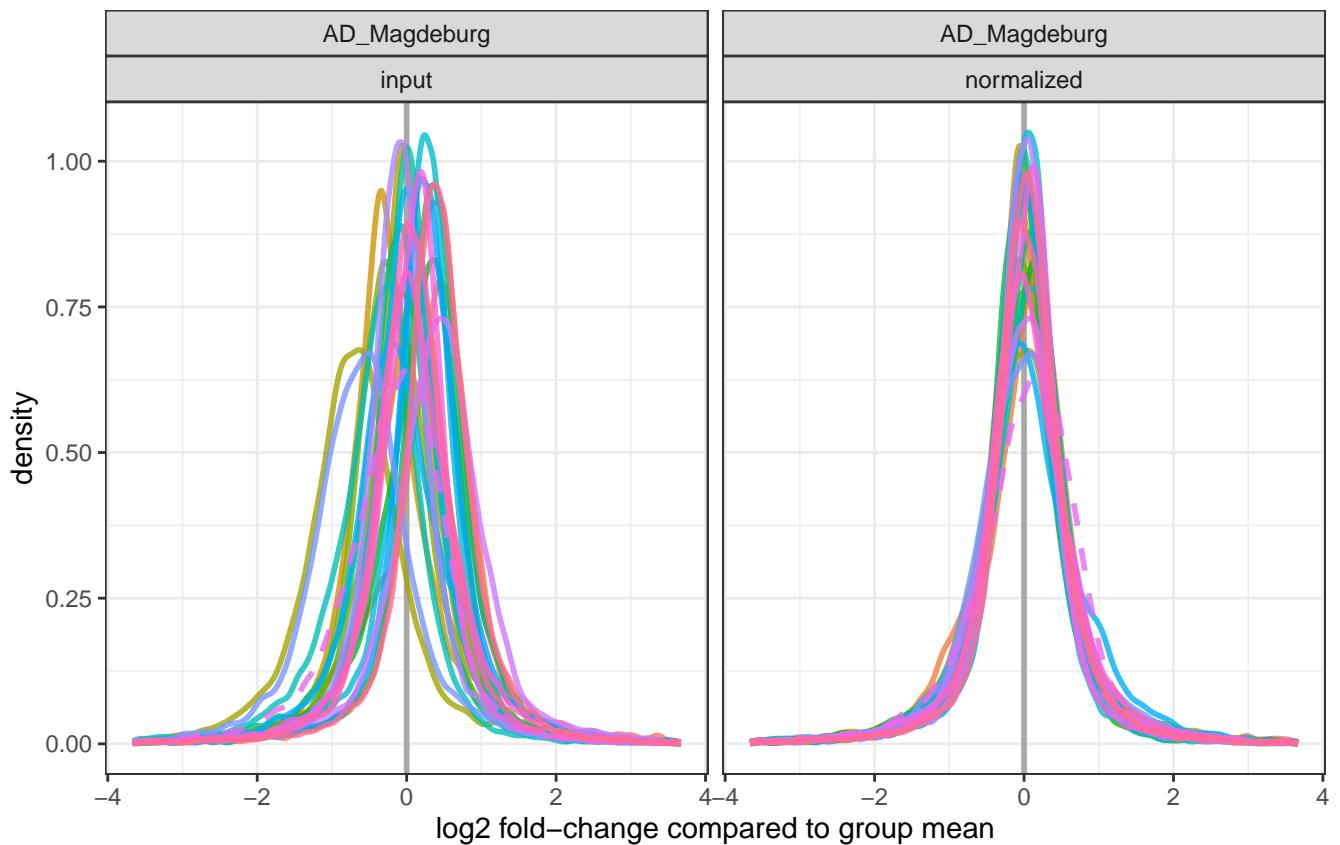
## 1.6 variation among replicates

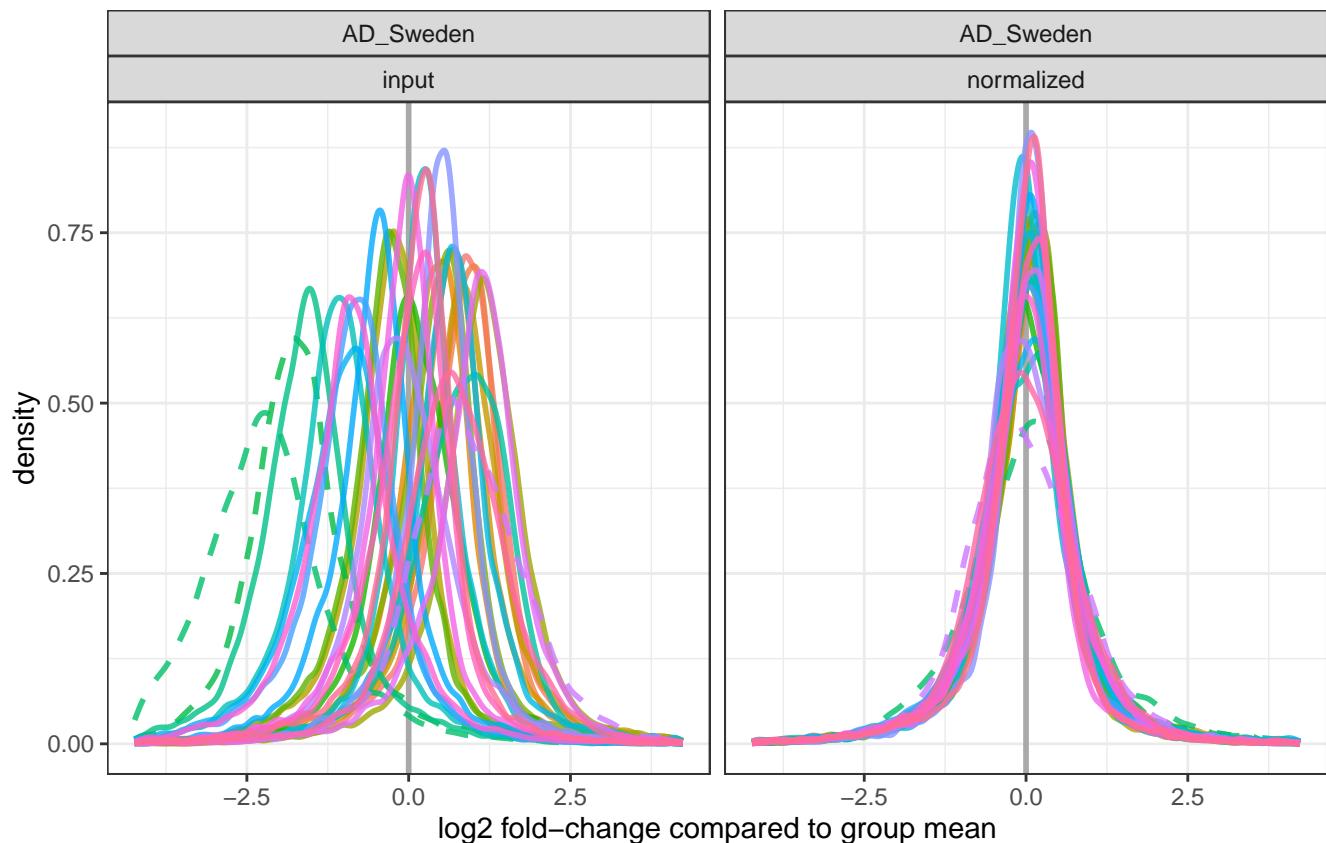
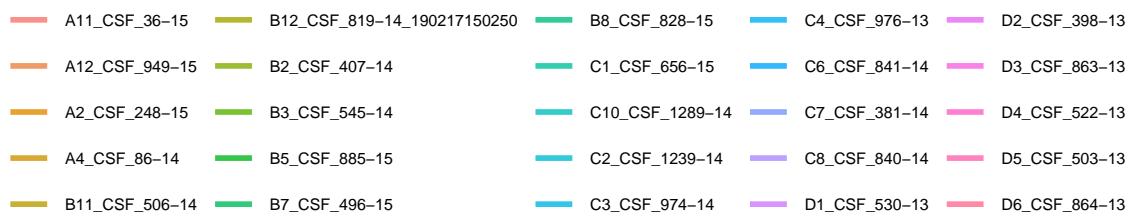
The reproducibility of replicate measurements is expressed in three different analyses. First, the difference between peptide intensities in each sample are compared to the mean value among all replicates (foldchange distributions). Next, the Coefficient of Variation (CoV) is used as a metric for reproducibility to explore how much the CoV within a sample group can be improved by removing a single sample (eg; if CoV strongly improved after removing sample s, it could be regarded as an outlier). Finally, the CoV within each sample group is visualized as a boxplot and a violin plot, figures commonly seen in proteomics literature and useful for comparing across experiments (of similar protocol).

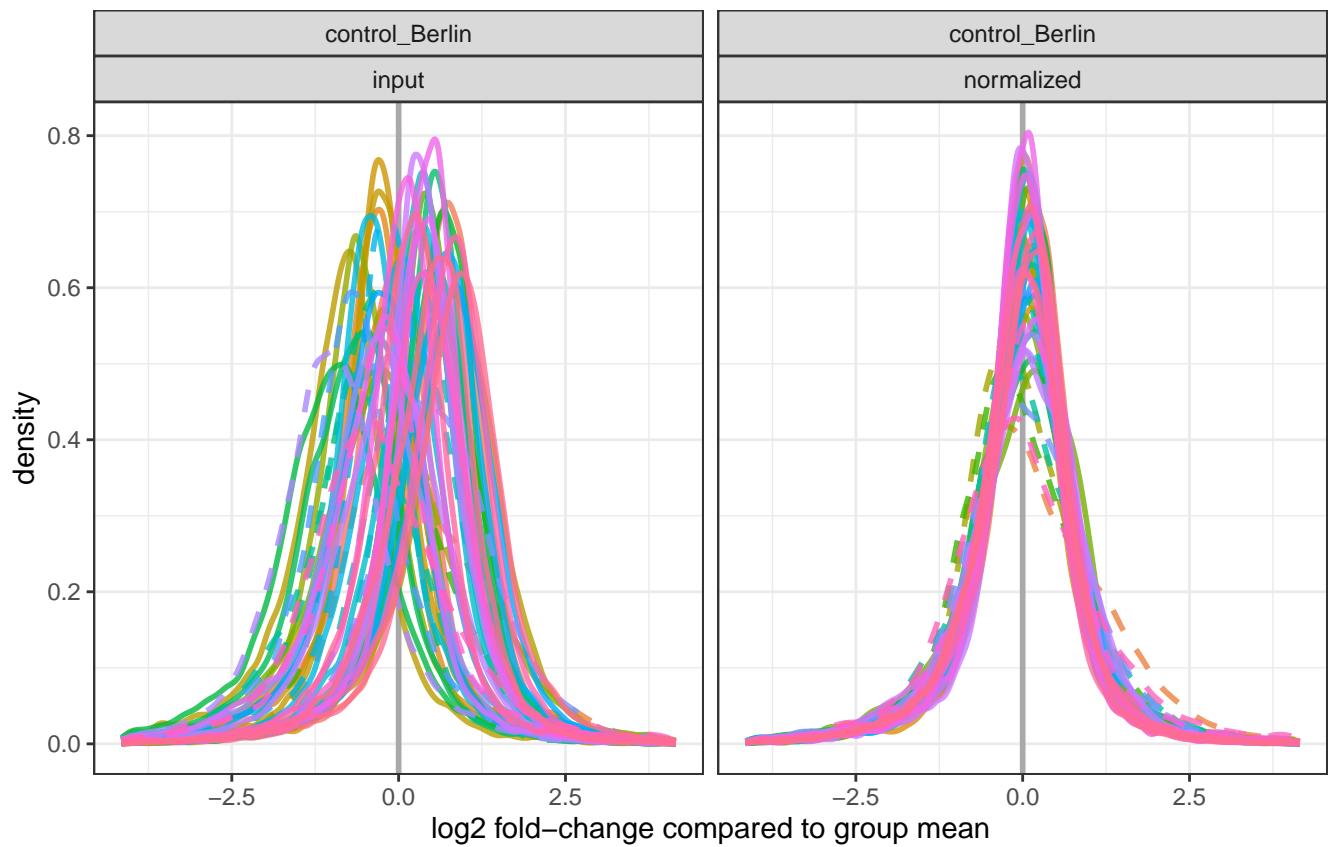
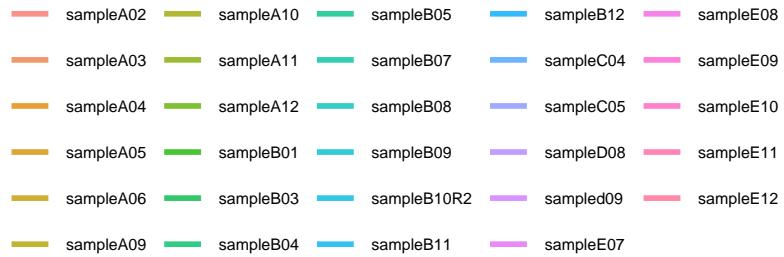
### 1.6.1 within-group foldchange distributions

The foldchange of all peptides in a sample is compared to their respective mean value over all samples in the group. This visualizes how strongly each sample deviates from other samples in the same group which helps identify outlier samples. The same data was used as detailed in the “retention time” section above. Samples marked as ‘exclude’ in the provided sample metadata table are visualized as dashed lines.

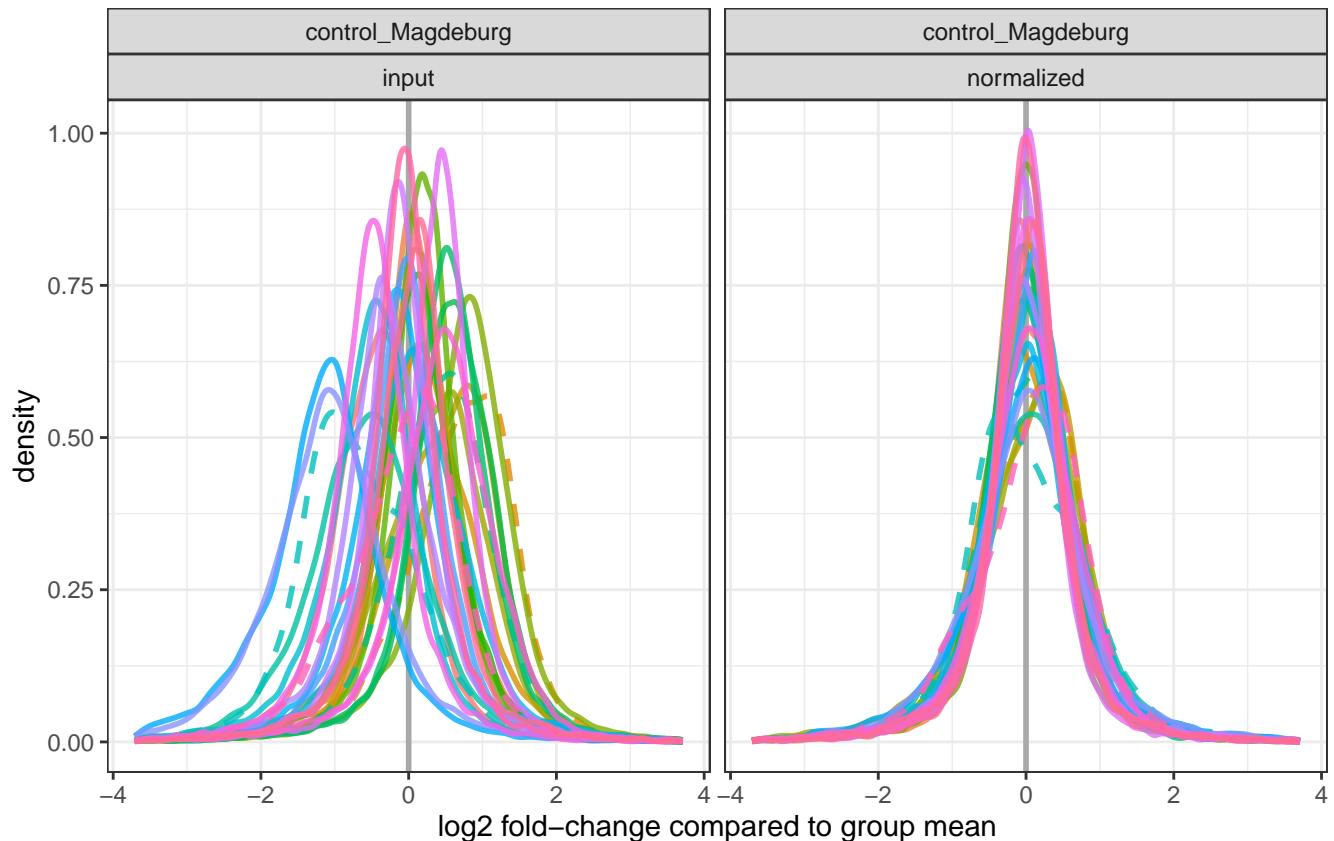


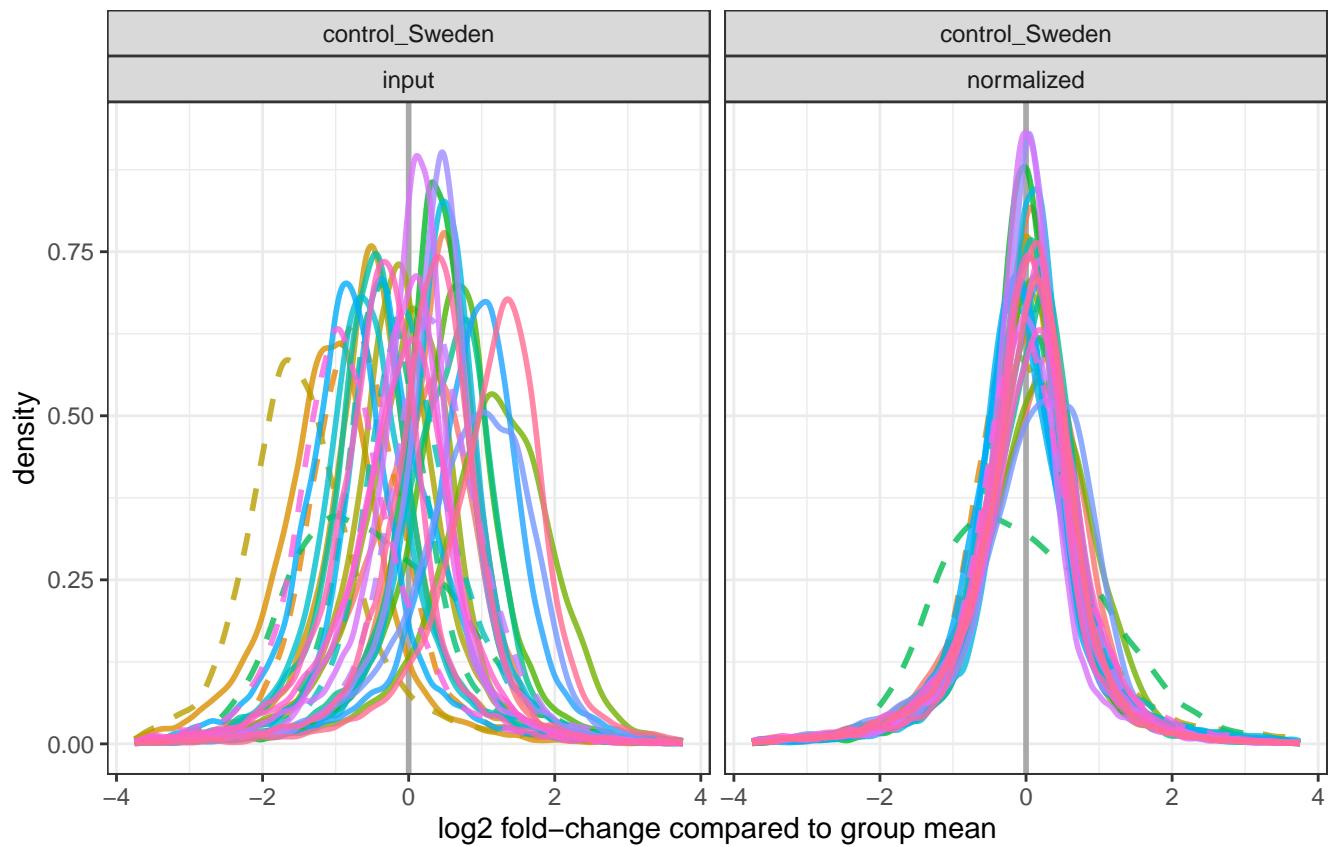
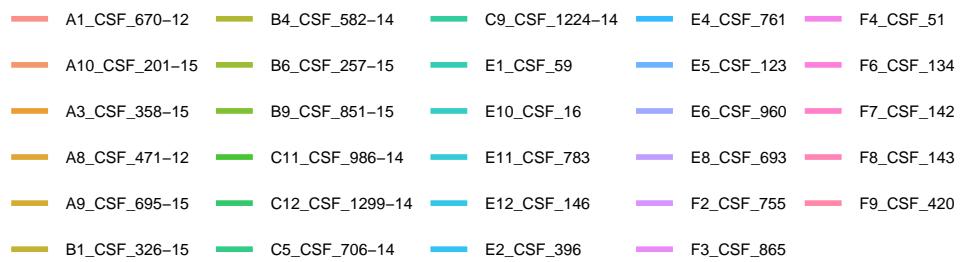


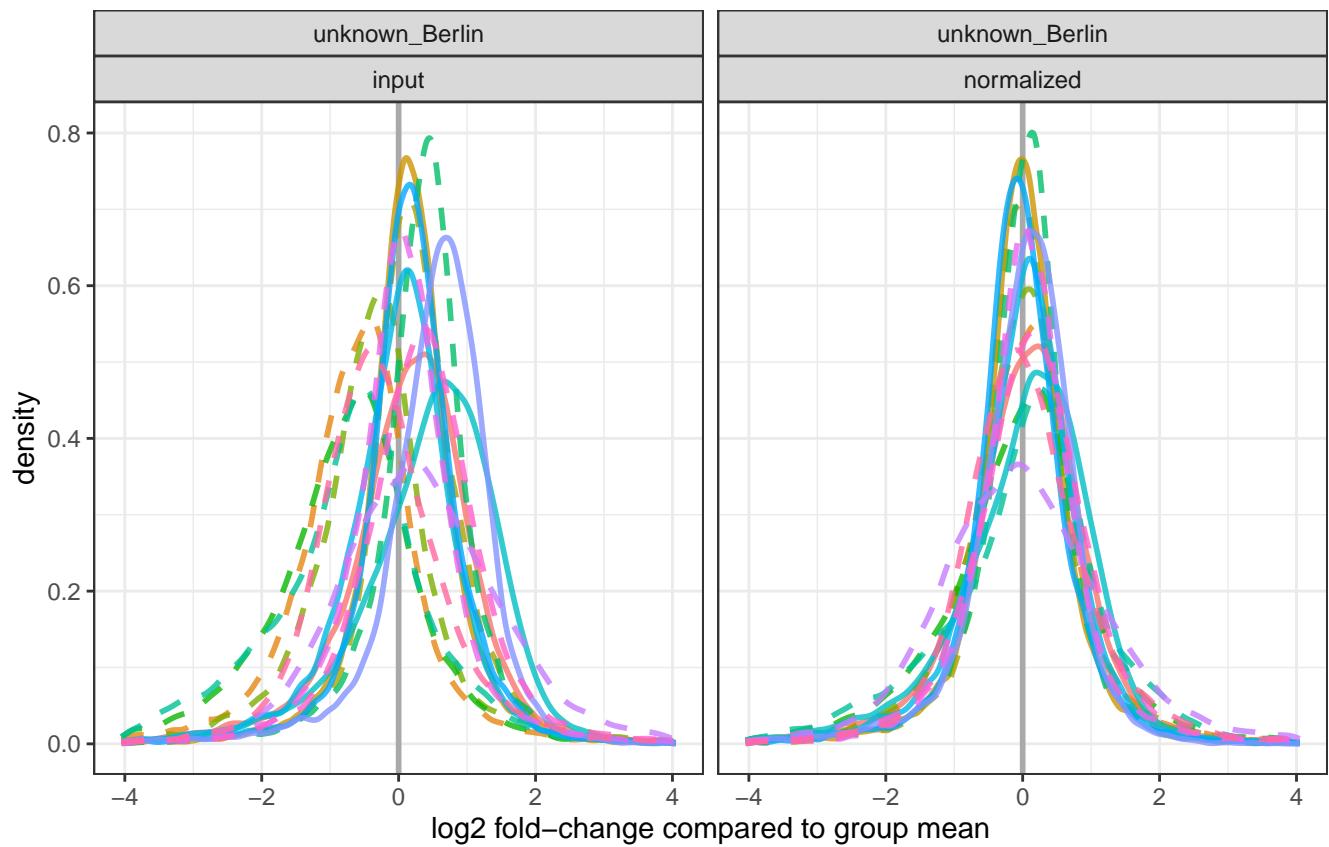
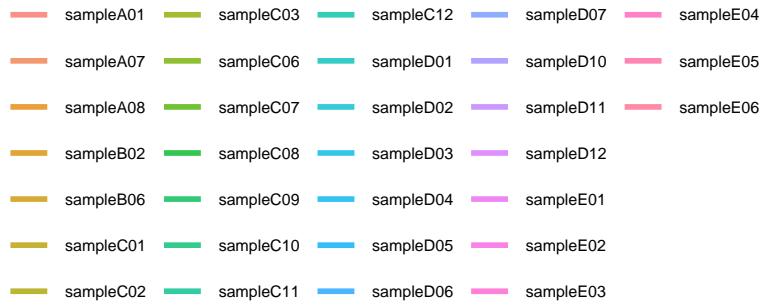


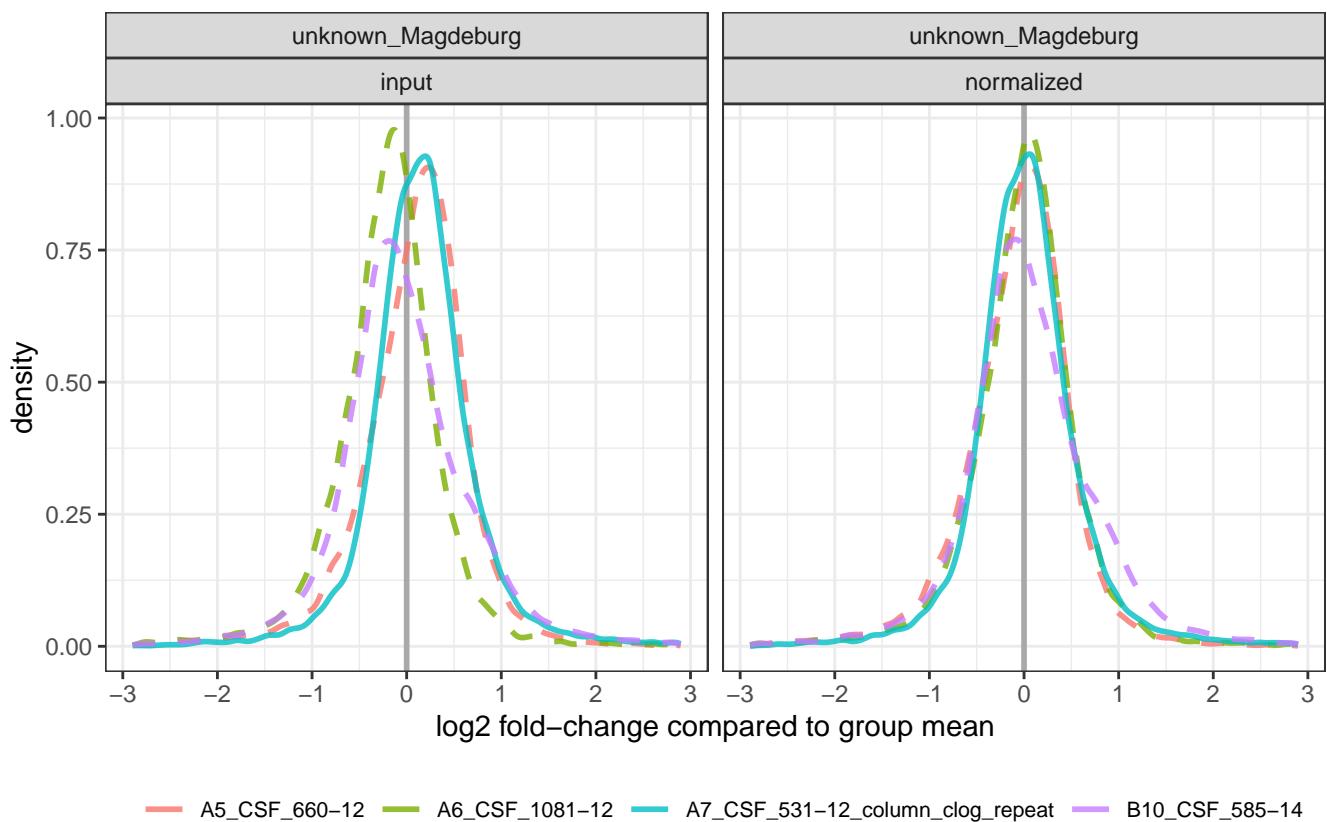


A10_82876065	B8_79116161	D12_168744796_repeat2	E7_81758548	G11_79112692_repeat
A11_129183842	B9_79059821	D6_129182306_repeat	E8_129188551_190304214129	G12_190225664
A12_190103450	C10_79069292	D7_81784062	E9_82882320_repeat	G6_168744541
A5_81790643	C11_129190335	D8_129182061	F10_190102946_repeat	G7_79061594
A7_79061681_repeat	C12_168743858_repeat	D9_82885545_repeat	F12_190100162	G8_82872412
A8_79114549_repeat	C5_81760601	E10_190225272	F5_129184099_repeat	G9_82879903_1903121815
A9_82544855	C6_82881358	E11_82871132_test_190306003105	F6_129189909_190308011553	H11_190103151
B10_79069372	C7_81790598_repeat2	E12_190226223	F7_81784602_repeat	H5_82544762_190309202C
B12_82872039	C9_79114900	E5_129187004_repeat	F8_190224717_repeat	H7_129187195
B6_79113519	D10_190224386	E6_129188085	G10_190102976	H9_82875993



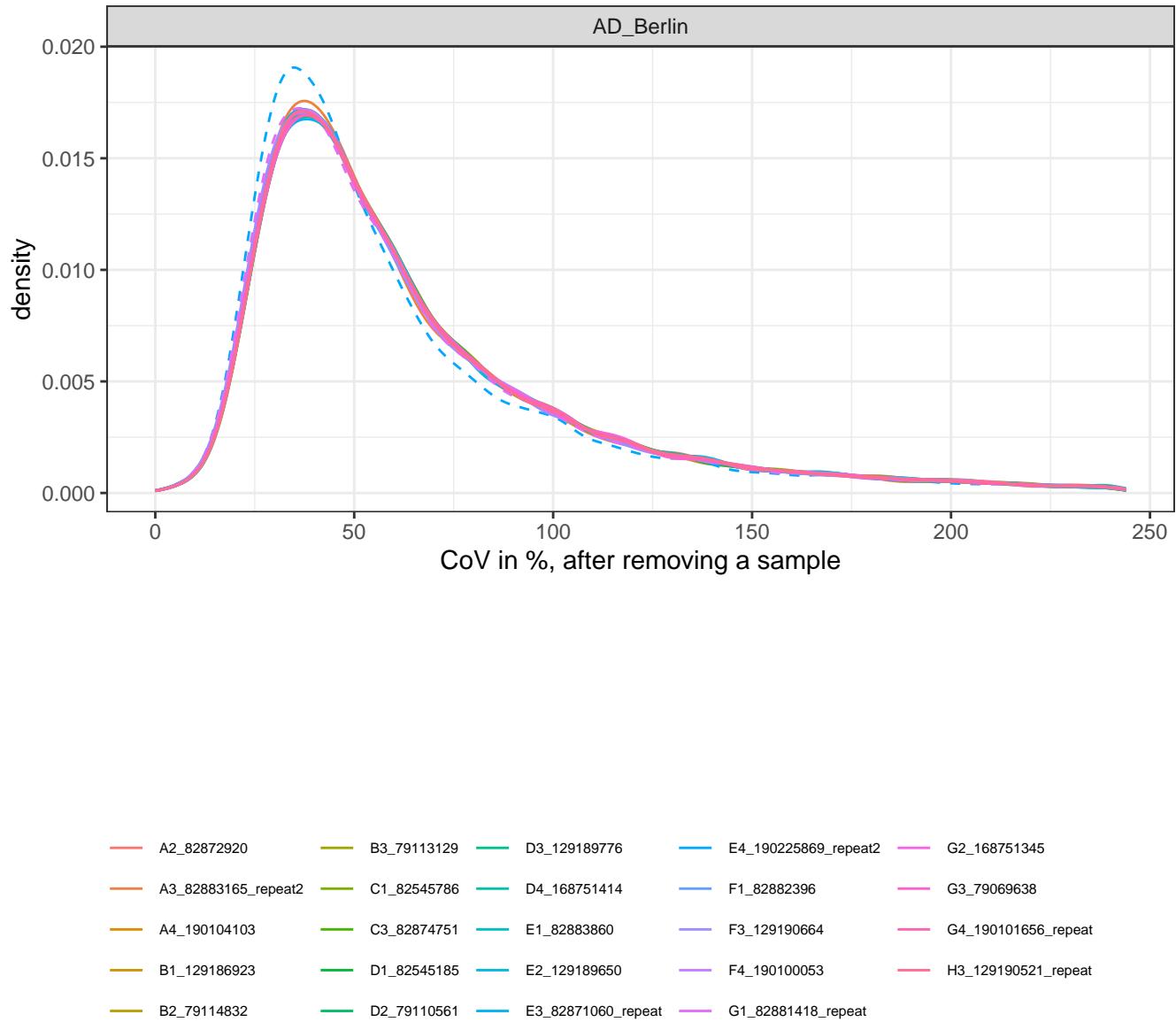


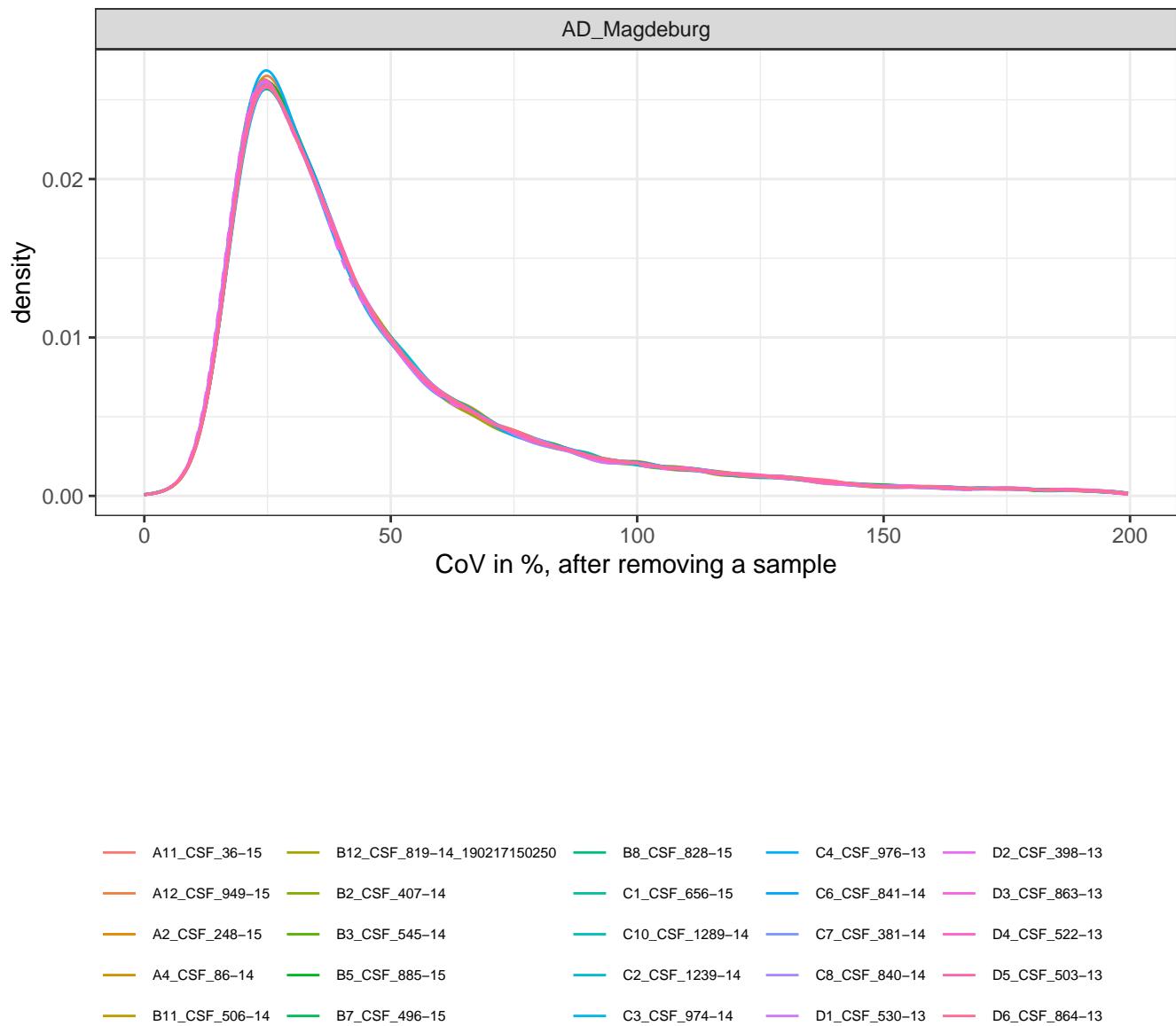


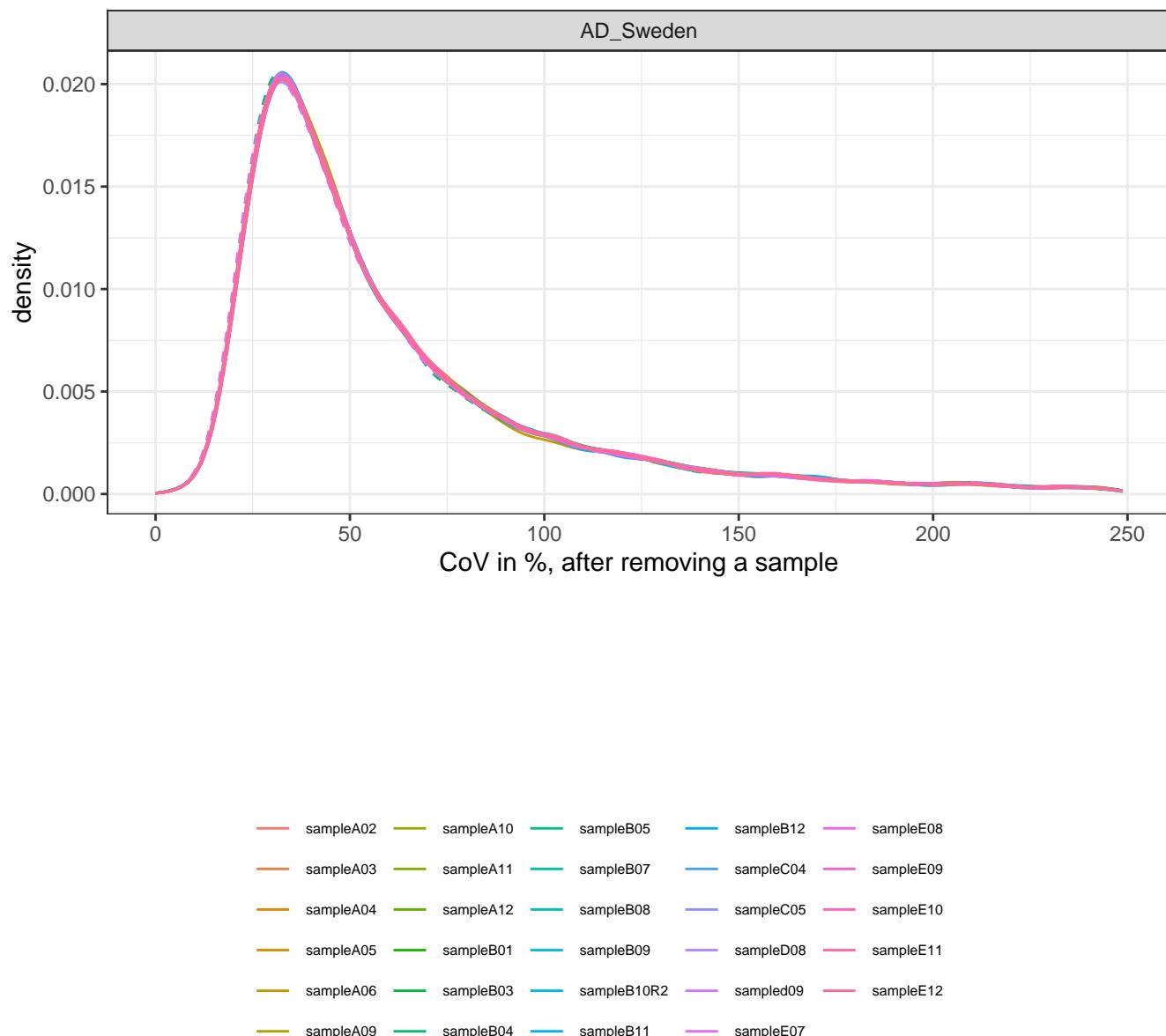


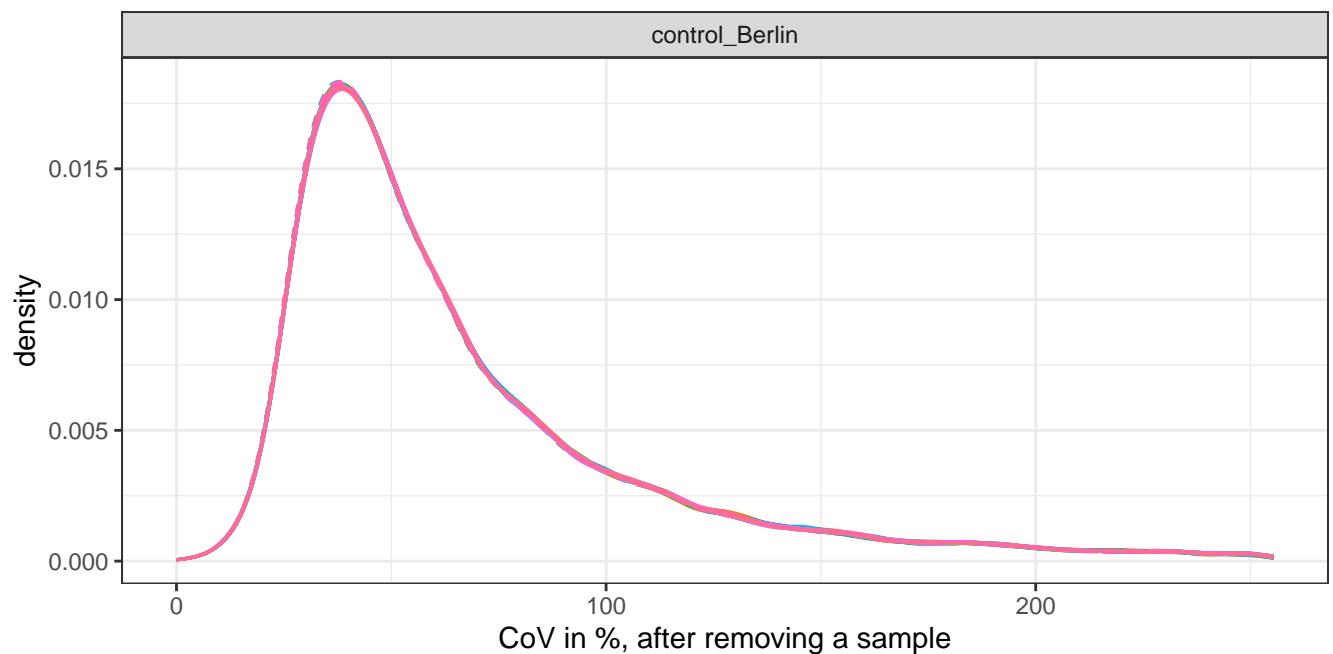
### 1.6.2 CoV, leave-one-out

The figures below describe the effect of removing a particular sample prior to within-group Coefficient of Variation (CoV) computation. The lower the CoV distribution is for a sample, the better reproducibility we get by excluding it. Only sample groups with at least 4 replicates can be used for this analysis, so 3 samples remain after leaving one out. Samples marked as ‘exclude’ in the provided sample metadata are included in these analyses (shown as dashed lines), and only peptides with at least 3 data points across replicate samples (after leave-one-out) are used for each CoV computation.

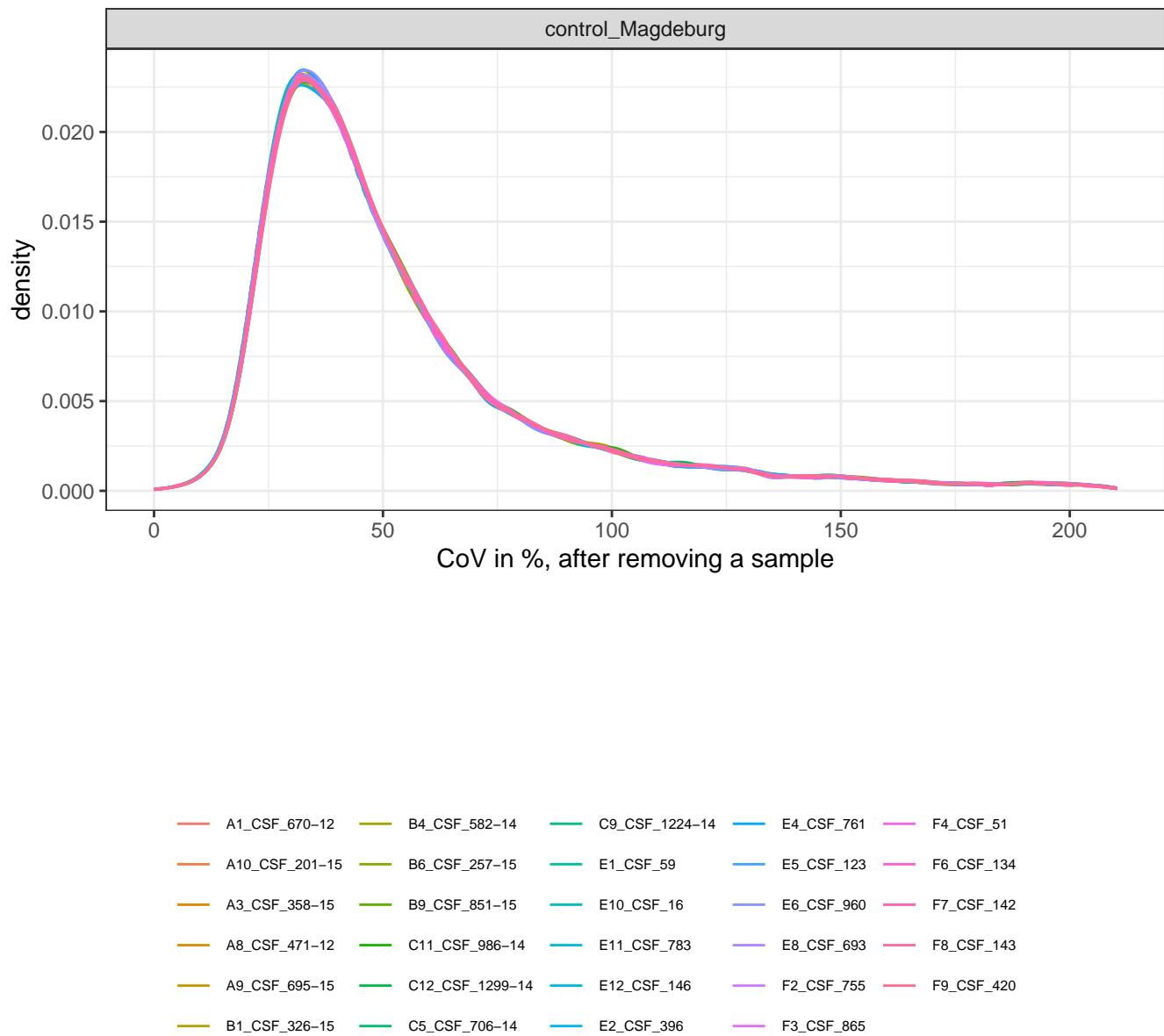


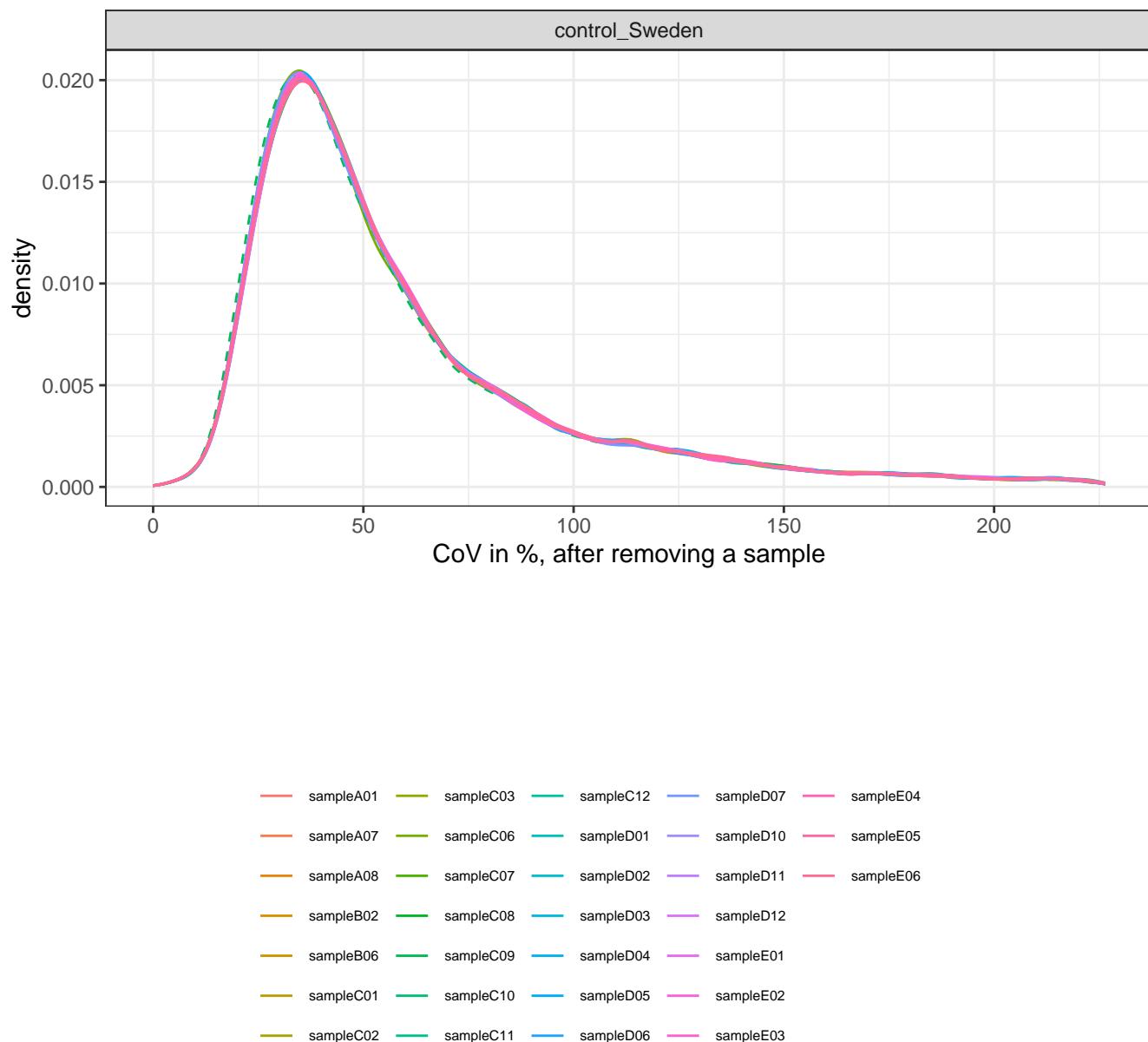


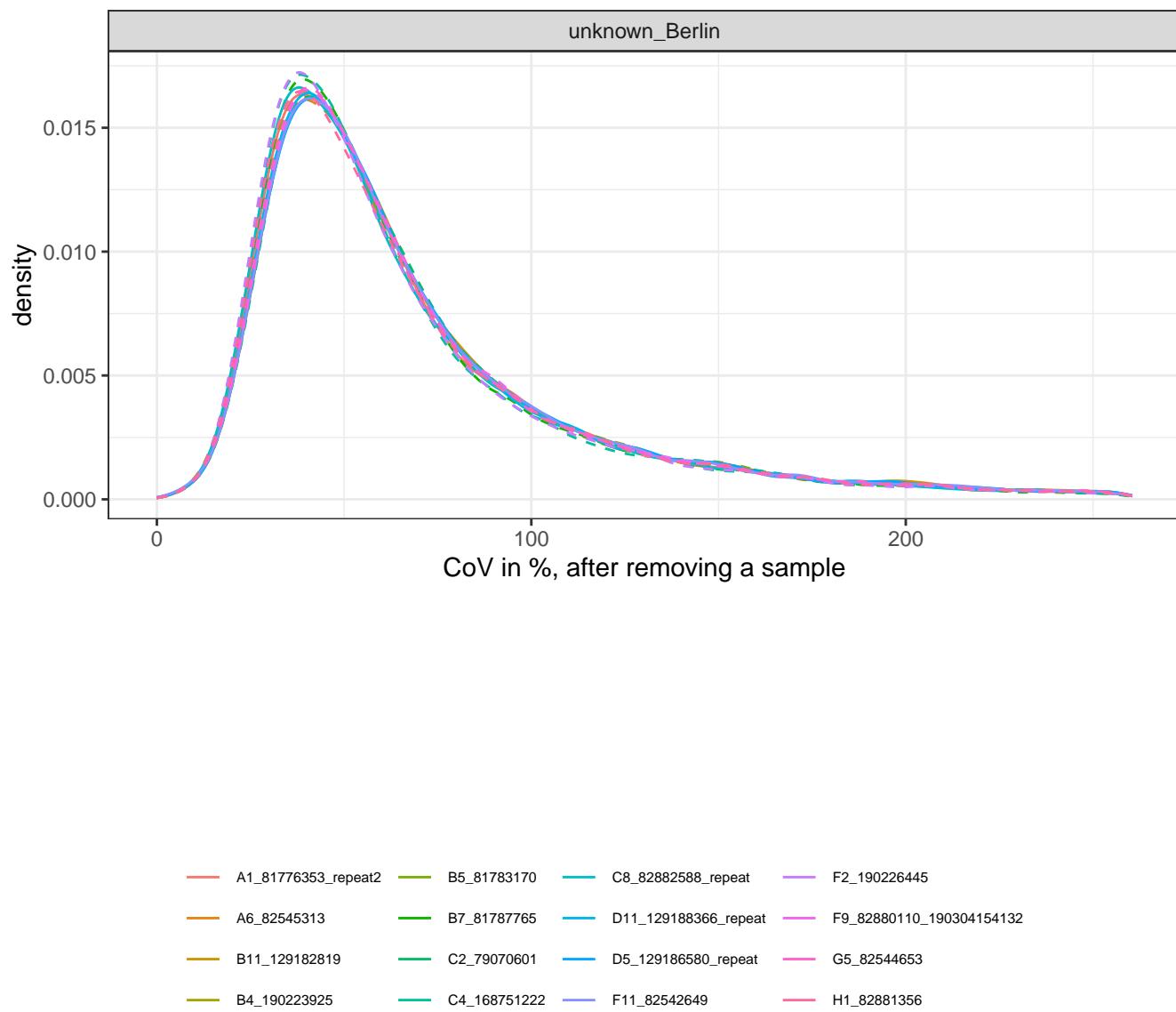


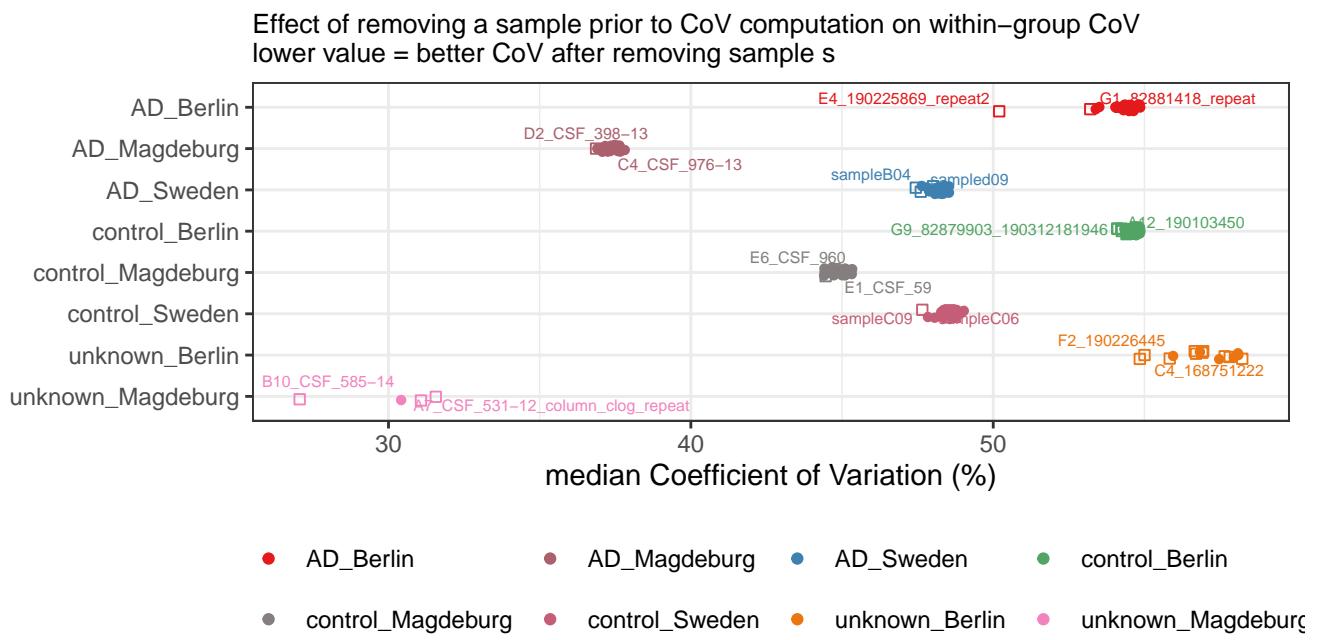
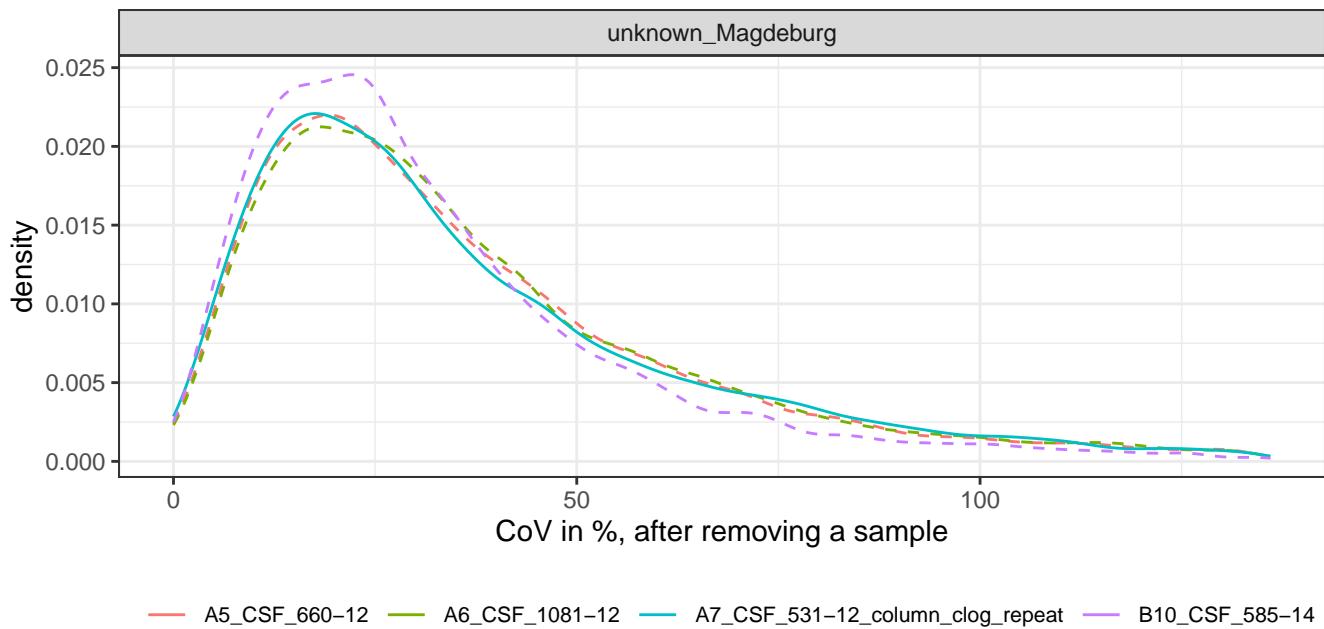


A10_82876065	B8_79116161	D12_168744796_repeat2	E7_81758548	G11_79112692_repeat
A11_129183842	B9_79059821	D6_129182306_repeat	E8_129188551_190304214129	G12_190225664
A12_190103450	C10_79069292	D7_81784062	E9_82882320_repeat	G6_168744541
A5_81790643	C11_129190335	D8_129182061	F10_190102946_repeat	G7_79061594
A7_79061681_repeat	C12_168743858_repeat	D9_82885545_repeat	F12_190100162	G8_82872412
A8_79114549_repeat	C5_81760601	E10_190225272	F5_129184099_repeat	G9_82879903_190312181
A9_82544855	C6_82881358	E11_82871132_test_190306003105	F6_129189909_190308011553	H11_190103151
B10_79069372	C7_81790598_repeat2	E12_190226223	F7_81784602_repeat	H5_82544762_190309202
B12_82872039	C9_79114900	E5_129187004_repeat	F8_190224717_repeat	H7_129187195
B6_79113519	D10_190224386	E6_129188085	G10_190102976	H9_82875993









shortname	group	exclude	median	CoV
E4_190225869_repeat2	AD_Berlin	TRUE	50.2	
G1_82881418_repeat	AD_Berlin	TRUE	53.2	
A3_82883165_repeat2	AD_Berlin	FALSE	53.4	
F3_129190664	AD_Berlin	FALSE	53.5	
E2_129189650	AD_Berlin	FALSE	54.0	
B2_79114832	AD_Berlin	FALSE	54.1	
D1_82545185	AD_Berlin	FALSE	54.1	
A2_82872920	AD_Berlin	FALSE	54.2	
G3_79069638	AD_Berlin	FALSE	54.3	
D3_129189776	AD_Berlin	FALSE	54.3	
H3_129190521_repeat	AD_Berlin	FALSE	54.3	
E1_82883860	AD_Berlin	FALSE	54.4	
F4_190100053	AD_Berlin	FALSE	54.5	
F1_82882396	AD_Berlin	FALSE	54.5	
C1_82545786	AD_Berlin	FALSE	54.5	
E3_82871060_repeat	AD_Berlin	FALSE	54.6	
G4_190101656_repeat	AD_Berlin	FALSE	54.6	
B1_129186923	AD_Berlin	FALSE	54.6	
D4_168751414	AD_Berlin	FALSE	54.6	
G2_168751345	AD_Berlin	FALSE	54.8	
D2_79110561	AD_Berlin	FALSE	54.8	
A4_190104103	AD_Berlin	FALSE	54.8	
C3_82874751	AD_Berlin	FALSE	54.8	
B3_79113129	AD_Berlin	FALSE	54.9	
D2_CSF_398-13	AD_Magdeburg	TRUE	36.9	
C4_CSF_976-13	AD_Magdeburg	FALSE	36.9	
B12_CSF_819-14_190217150250	AD_Magdeburg	FALSE	37.1	
C7_CSF_381-14	AD_Magdeburg	FALSE	37.1	
A12_CSF_949-15	AD_Magdeburg	FALSE	37.2	
A2_CSF_248-15	AD_Magdeburg	FALSE	37.2	
D1_CSF_530-13	AD_Magdeburg	FALSE	37.2	
C10_CSF_1289-14	AD_Magdeburg	FALSE	37.3	
D4_CSF_522-13	AD_Magdeburg	FALSE	37.3	
B3_CSF_545-14	AD_Magdeburg	FALSE	37.3	
A11_CSF_36-15	AD_Magdeburg	FALSE	37.4	
B5_CSF_885-15	AD_Magdeburg	FALSE	37.4	
B8_CSF_828-15	AD_Magdeburg	FALSE	37.5	
B2_CSF_407-14	AD_Magdeburg	FALSE	37.5	
A4_CSF_86-14	AD_Magdeburg	FALSE	37.5	
B11_CSF_506-14	AD_Magdeburg	FALSE	37.6	
D5_CSF_503-13	AD_Magdeburg	FALSE	37.6	
C3_CSF_974-14	AD_Magdeburg	FALSE	37.6	
B7_CSF_496-15	AD_Magdeburg	FALSE	37.6	
C6_CSF_841-14	AD_Magdeburg	FALSE	37.6	
C8_CSF_840-14	AD_Magdeburg	FALSE	37.7	
D3_CSF_863-13	AD_Magdeburg	FALSE	37.7	
C1_CSF_656-15	AD_Magdeburg	FALSE	37.7	
C2_CSF_1239-14	AD_Magdeburg	FALSE	37.8	
D6_CSF_864-13	AD_Magdeburg	FALSE	37.8	
sampleB04	AD_Sweden	TRUE	47.4	

shortname	group	exclude	median	CoV
sampled09	AD_Sweden	TRUE	47.6	
sampleA06	AD_Sweden	FALSE	47.6	
sampleB11	AD_Sweden	FALSE	47.8	
sampleE09	AD_Sweden	FALSE	47.9	
sampleE11	AD_Sweden	FALSE	47.9	
sampleB03	AD_Sweden	TRUE	48.0	
sampleB01	AD_Sweden	FALSE	48.0	
sampleB10R2	AD_Sweden	FALSE	48.1	
sampleB07	AD_Sweden	FALSE	48.1	
sampleD08	AD_Sweden	FALSE	48.1	
sampleC04	AD_Sweden	FALSE	48.1	
sampleA03	AD_Sweden	FALSE	48.2	
sampleB08	AD_Sweden	FALSE	48.2	
sampleA10	AD_Sweden	FALSE	48.2	
sampleB05	AD_Sweden	FALSE	48.2	
sampleE07	AD_Sweden	FALSE	48.2	
sampleA12	AD_Sweden	FALSE	48.3	
sampleA05	AD_Sweden	FALSE	48.3	
sampleA11	AD_Sweden	FALSE	48.3	
sampleB12	AD_Sweden	FALSE	48.3	
sampleA04	AD_Sweden	FALSE	48.4	
sampleE10	AD_Sweden	FALSE	48.4	
sampleE08	AD_Sweden	FALSE	48.4	
sampleA02	AD_Sweden	FALSE	48.4	
sampleA09	AD_Sweden	FALSE	48.4	
sampleE12	AD_Sweden	FALSE	48.5	
sampleB09	AD_Sweden	FALSE	48.5	
sampleC05	AD_Sweden	FALSE	48.5	
G9_82879903_190312181946	control_Berlin	TRUE	54.1	
A12_190103450	control_Berlin	TRUE	54.1	
C5_81760601	control_Berlin	TRUE	54.2	
D7_81784062	control_Berlin	TRUE	54.3	
G7_79061594	control_Berlin	TRUE	54.3	
C7_81790598_repeat2	control_Berlin	FALSE	54.3	
A8_79114549_repeat	control_Berlin	FALSE	54.4	
F7_81784602_repeat	control_Berlin	FALSE	54.4	
F12_190100162	control_Berlin	TRUE	54.4	
B8_79116161	control_Berlin	TRUE	54.5	
G10_190102976	control_Berlin	FALSE	54.5	
C11_129190335	control_Berlin	FALSE	54.5	
F10_190102946_repeat	control_Berlin	TRUE	54.5	
F6_129189909_190308011553	control_Berlin	TRUE	54.5	
F5_129184099_repeat	control_Berlin	FALSE	54.6	
C9_79114900	control_Berlin	FALSE	54.6	
E10_190225272	control_Berlin	TRUE	54.6	
B6_79113519	control_Berlin	FALSE	54.6	
G8_82872412	control_Berlin	FALSE	54.6	
E8_129188551_190304214129	control_Berlin	FALSE	54.6	
H11_190103151	control_Berlin	FALSE	54.6	
D12_168744796_repeat2	control_Berlin	TRUE	54.6	

shortname	group	exclude	median	CoV
H9_82875993	control_Berlin	FALSE	54.6	
B12_82872039	control_Berlin	FALSE	54.6	
A9_82544855	control_Berlin	FALSE	54.6	
E7_81758548	control_Berlin	FALSE	54.6	
C10_79069292	control_Berlin	FALSE	54.7	
B9_79059821	control_Berlin	FALSE	54.7	
E11_82871132_test_190306003105	control_Berlin	FALSE	54.7	
A10_82876065	control_Berlin	FALSE	54.7	
E6_129188085	control_Berlin	TRUE	54.7	
D8_129182061	control_Berlin	TRUE	54.7	
A5_81790643	control_Berlin	FALSE	54.7	
E5_129187004_repeat	control_Berlin	FALSE	54.7	
E12_190226223	control_Berlin	FALSE	54.7	
E9_82882320_repeat	control_Berlin	FALSE	54.7	
C6_82881358	control_Berlin	FALSE	54.7	
H7_129187195	control_Berlin	FALSE	54.7	
D6_129182306_repeat	control_Berlin	FALSE	54.7	
D10_190224386	control_Berlin	FALSE	54.8	
B10_79069372	control_Berlin	FALSE	54.8	
G6_168744541	control_Berlin	FALSE	54.8	
G12_190225664	control_Berlin	FALSE	54.8	
A11_129183842	control_Berlin	FALSE	54.8	
H5_82544762_190309202008	control_Berlin	FALSE	54.8	
C12_168743858_repeat	control_Berlin	FALSE	54.8	
A7_79061681_repeat	control_Berlin	FALSE	54.8	
D9_82885545_repeat	control_Berlin	FALSE	54.9	
G11_79112692_repeat	control_Berlin	FALSE	54.9	
F8_190224717_repeat	control_Berlin	FALSE	54.9	
E6_CSF_960	control_Magdeburg	FALSE	44.4	
E1_CSF_59	control_Magdeburg	FALSE	44.4	
E10_CSF_16	control_Magdeburg	TRUE	44.5	
B4_CSF_582-14	control_Magdeburg	FALSE	44.5	
F7_CSF_142	control_Magdeburg	TRUE	44.5	
E4_CSF_761	control_Magdeburg	FALSE	44.5	
C9_CSF_1224-14	control_Magdeburg	TRUE	44.7	
A9_CSF_695-15	control_Magdeburg	FALSE	44.7	
E8_CSF_693	control_Magdeburg	FALSE	44.7	
A3_CSF_358-15	control_Magdeburg	TRUE	44.8	
E12_CSF_146	control_Magdeburg	FALSE	44.8	
B6_CSF_257-15	control_Magdeburg	FALSE	44.9	
E2_CSF_396	control_Magdeburg	FALSE	44.9	
F6_CSF_134	control_Magdeburg	FALSE	45.0	
A1_CSF_670-12	control_Magdeburg	FALSE	45.0	
A8_CSF_471-12	control_Magdeburg	FALSE	45.0	
C11_CSF_986-14	control_Magdeburg	FALSE	45.0	
E5_CSF_123	control_Magdeburg	FALSE	45.0	
E11_CSF_783	control_Magdeburg	FALSE	45.1	
C12_CSF_1299-14	control_Magdeburg	FALSE	45.1	
F9_CSF_420	control_Magdeburg	FALSE	45.1	
A10_CSF_201-15	control_Magdeburg	FALSE	45.1	

shortname	group	exclude	median CoV
F2_CSF_755	control_Magdeburg	FALSE	45.1
B1_CSF_326-15	control_Magdeburg	FALSE	45.1
F4_CSF_51	control_Magdeburg	FALSE	45.1
C5_CSF_706-14	control_Magdeburg	FALSE	45.2
F8_CSF_143	control_Magdeburg	FALSE	45.3
F3_CSF_865	control_Magdeburg	FALSE	45.3
B9_CSF_851-15	control_Magdeburg	FALSE	45.3
sampleC09	control_Sweden	TRUE	47.7
sampleC06	control_Sweden	FALSE	47.8
sampleD07	control_Sweden	FALSE	48.1
sampleA01	control_Sweden	FALSE	48.2
sampleB02	control_Sweden	FALSE	48.3
sampleE03	control_Sweden	FALSE	48.4
sampleD02	control_Sweden	FALSE	48.4
sampleD05	control_Sweden	FALSE	48.5
sampleE02	control_Sweden	TRUE	48.5
sampleC01	control_Sweden	TRUE	48.6
sampleE01	control_Sweden	FALSE	48.6
sampleC02	control_Sweden	FALSE	48.6
sampleD04	control_Sweden	FALSE	48.6
sampleD01	control_Sweden	TRUE	48.6
sampleC07	control_Sweden	FALSE	48.6
sampleC11	control_Sweden	FALSE	48.6
sampleC03	control_Sweden	FALSE	48.6
sampleB06	control_Sweden	FALSE	48.7
sampleD06	control_Sweden	FALSE	48.7
sampleA08	control_Sweden	TRUE	48.7
sampleE06	control_Sweden	FALSE	48.7
sampleD11	control_Sweden	TRUE	48.7
sampleC10	control_Sweden	TRUE	48.7
sampleA07	control_Sweden	FALSE	48.7
sampleC12	control_Sweden	FALSE	48.7
sampleE04	control_Sweden	FALSE	48.8
sampleE05	control_Sweden	FALSE	48.8
sampleD10	control_Sweden	FALSE	48.8
sampleD03	control_Sweden	FALSE	48.8
sampleC08	control_Sweden	FALSE	48.9
sampleD12	control_Sweden	FALSE	49.0
F2_190226445	unknown_Berlin	TRUE	54.8
C4_168751222	unknown_Berlin	TRUE	55.0
B7_81787765	unknown_Berlin	TRUE	55.8
C8_82882588_repeat	unknown_Berlin	FALSE	55.9
B5_81783170	unknown_Berlin	TRUE	56.7
H1_82881356	unknown_Berlin	TRUE	56.7
A1_81776353_repeat2	unknown_Berlin	FALSE	56.8
G5_82544653	unknown_Berlin	TRUE	56.9
A6_82545313	unknown_Berlin	TRUE	56.9
D11_129188366_repeat	unknown_Berlin	FALSE	57.5
F9_82880110_190304154132	unknown_Berlin	TRUE	57.7
B4_190223925	unknown_Berlin	TRUE	57.8

shortname	group	exclude	median CoV
F11_82542649	unknown_Berlin	FALSE	58.0
D5_129186580_repeat	unknown_Berlin	FALSE	58.1
B11_129182819	unknown_Berlin	FALSE	58.1
C2_79070601	unknown_Berlin	TRUE	58.2
B10_CSF_585-14	unknown_Magdeburg	TRUE	27.1
A7_CSF_531-12_column_clog_repeat	unknown_Magdeburg	FALSE	30.4
A5_CSF_660-12	unknown_Magdeburg	TRUE	31.1
A6_CSF_1081-12	unknown_Magdeburg	TRUE	31.6

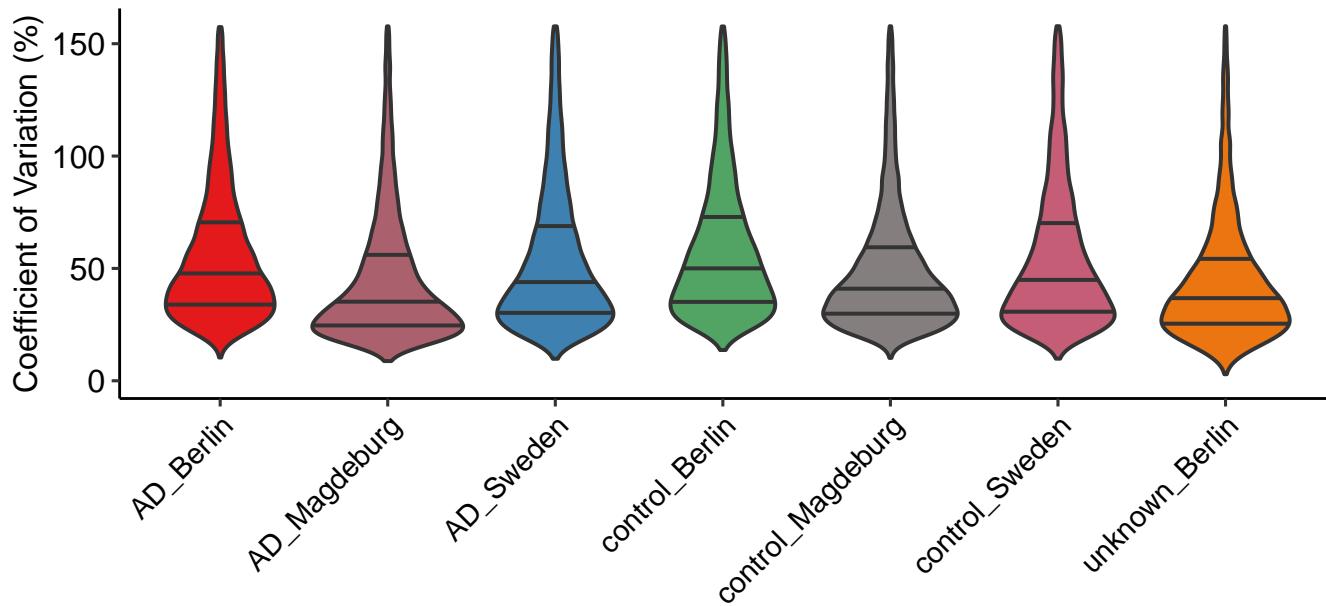
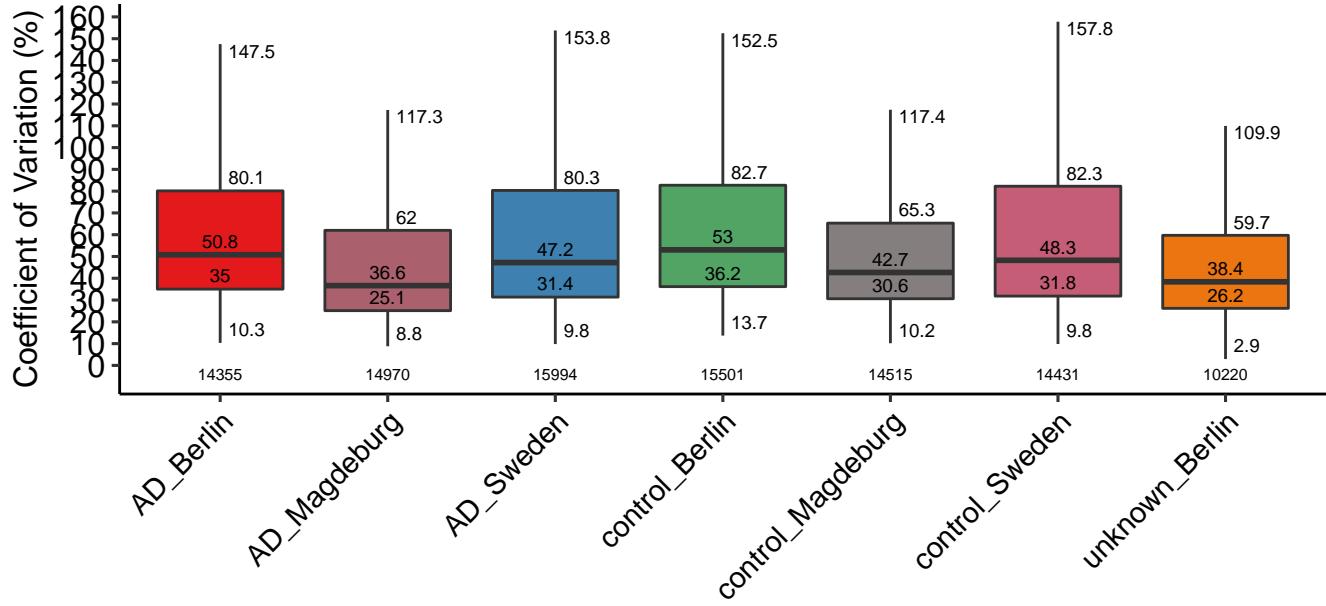
*Leave-one-out impact on within-group CoV (%)*

### 1.6.3 Coefficient of Variation

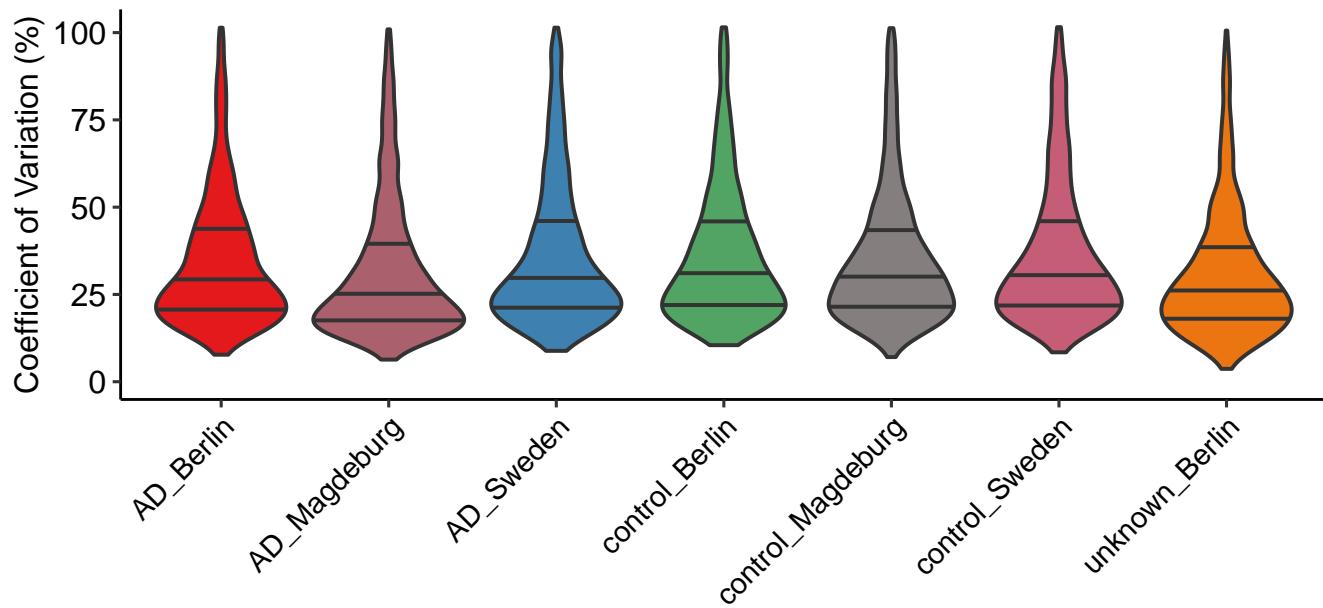
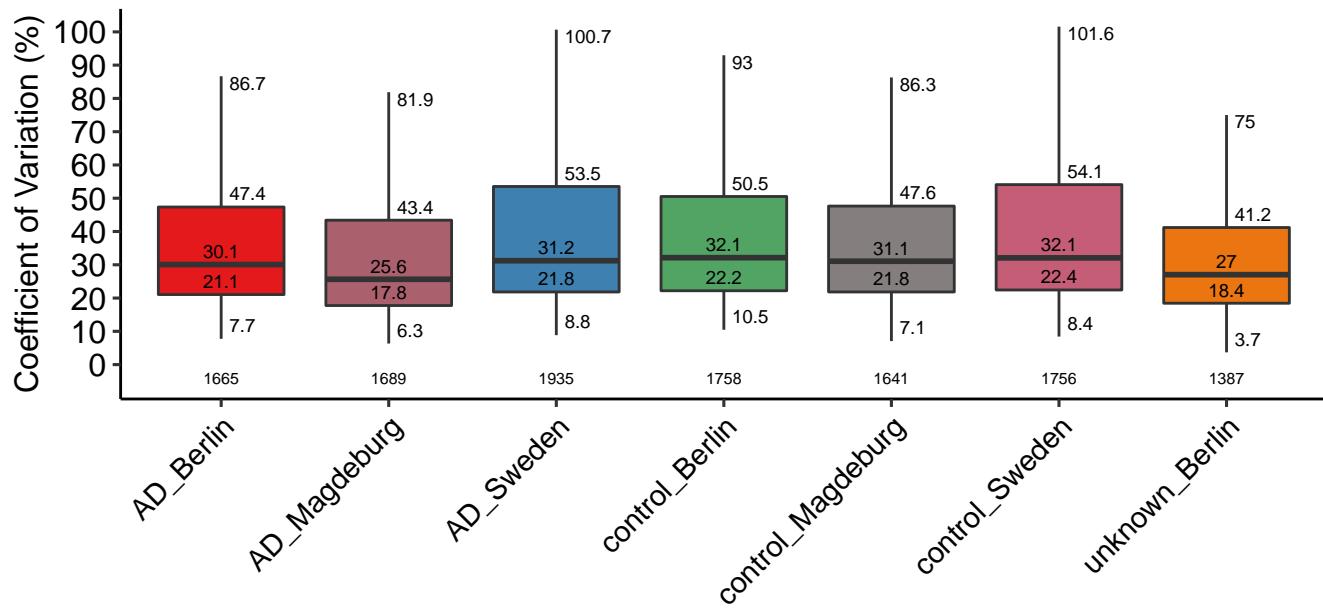
The Coefficient of Variation (CoV) is a quality metric for the reproducibility of replicate measurements, here visualized using box- and violin-plots.

Only samples that are NOT marked ‘exclude’ in the provided sample metadata and are in a sample group among at least 3 replicates are used for these figures. The user-specified filtering rules (eg; filter\_min\_detect, filter\_min\_peptide\_per\_prot, etc.) were applied within each sample group independently and remaining peptides were subsequently normalized. Only peptides with at least 3 data points across replicate samples are used for each CoV computation.

**Peptide-level CoV:**



**Protein-level CoV:** (analogous to peptide CoV's, but with additional MaxLFQ rollup to protein abundances)



## 1.7 PCA

A visualization of the first three PCA dimensions illustrates sample clustering. The goal of these figures is to detect global effects from a quality control perspective, such as samples from the same experiment batch clustering together, not to be sensitive to a minor subset of differentially abundant proteins (for which specialized statistical models can be applied downstream).

If additional sample metadata was provided, such as experiment batch, sample-prep dates, gel, etc., multiple PCA figures will be generated with respective color-codings. Users are encouraged to provide relevant experiment information as sample metadata and use these figures to search for unexpected batch effects.

The pcaMethods R package is used here to perform the Probabilistic PCA (PPCA). The set of peptides used for this analysis consists of those peptides that pass your filter criteria in every sample group. If any samples are marked as ‘exclude’ in the provided sample metadata, an additional PCA plot is generated with these samples included (depicting the ‘exclude’ samples as square symbols).

### Rationale behind data filter

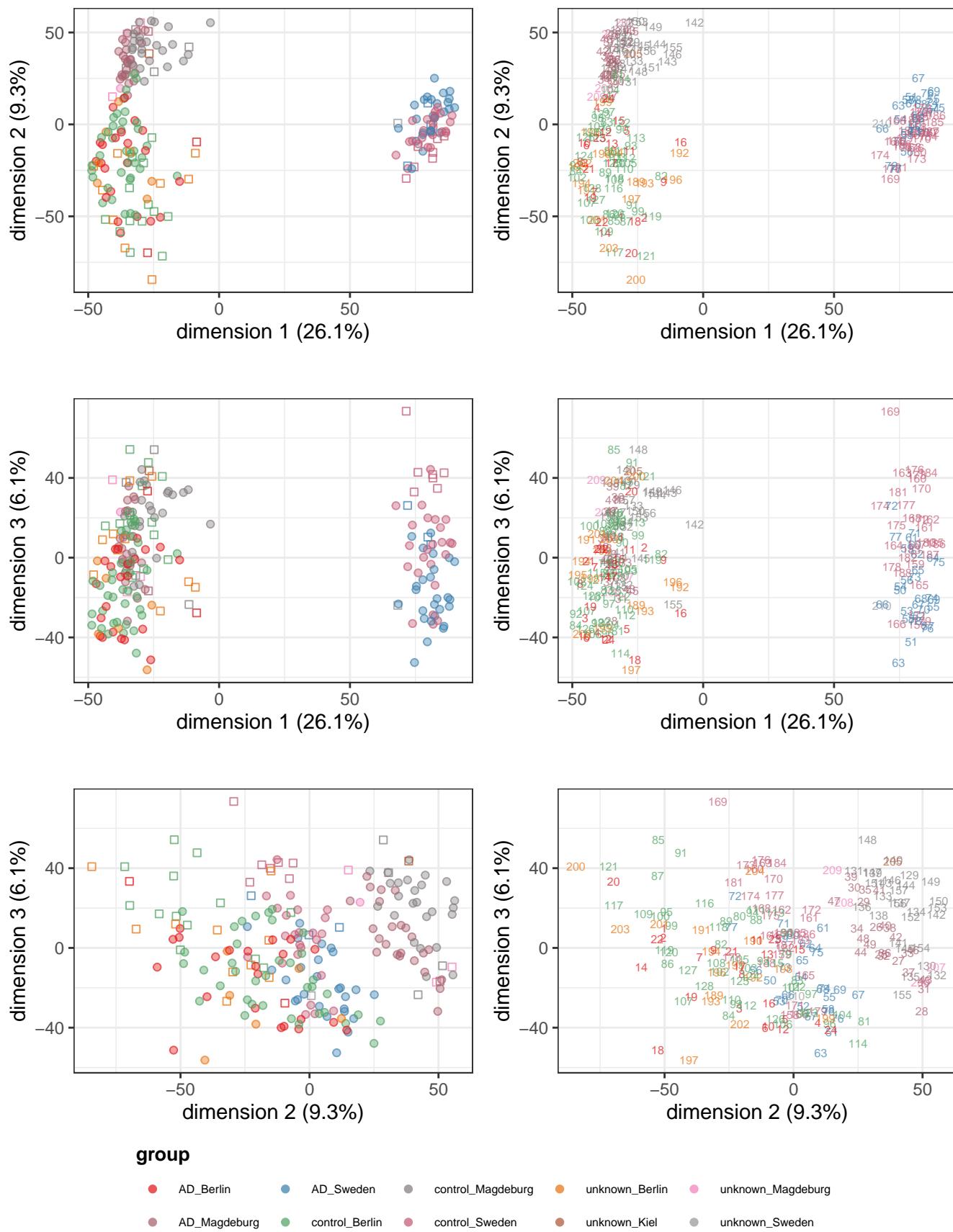
As mentioned above, the aim of the PCA figures is to identify global effects. To achieve this, we compute sample distances on the subset of peptides identified in each group which prevents rarely detected peptides/proteins from having a disproportionate effect on sample clustering. This pertains not only to ‘randomly detected contaminant proteins’ but also to proteins with abundance levels near the detection limit, which may be detected in only a subset of samples (eg; some measurements will be more successful/sensitive than others).

### Figure legends

The first 3 principle components compared visually (1 vs 2, 1 vs 3, 2 vs 3) on the rows. Left- and right-side panels on each row represent the same figure without and with sample labels. The principle components are shown on the axis labels together with their respective percentage of variance explained. Samples marked as ‘exclude’ in the provided sample metadata, if any, are visualized as square shapes.

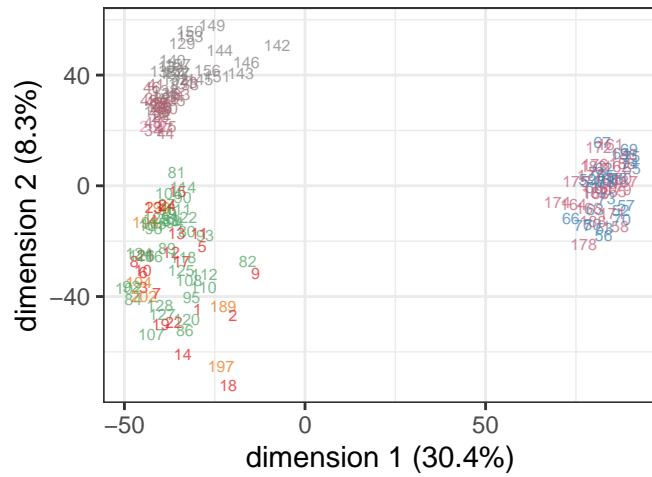
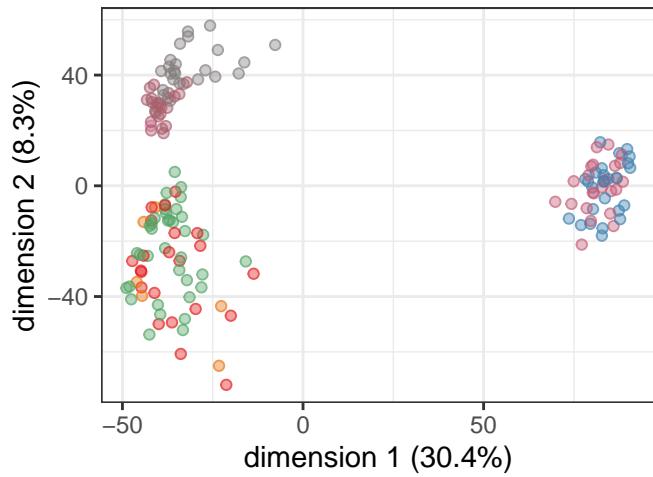
## PCA of all samples, including those flagged as 'exclude', using 8892 peptides

samples are labeled by "sample\_index" (described in samples.xlsx included with MS-DAP output)

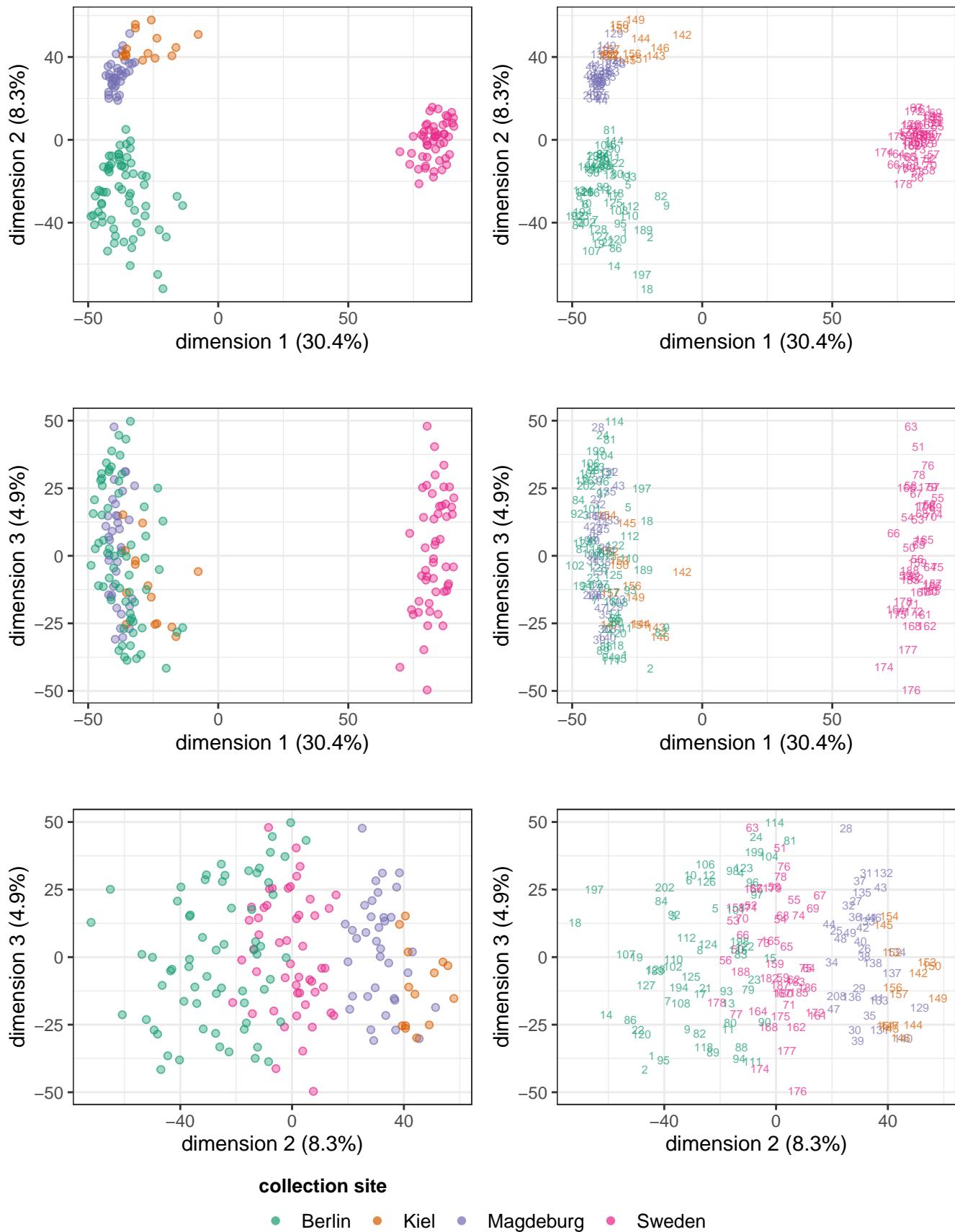


## PCA only on samples not flagged as 'exclude', using 9164 peptides

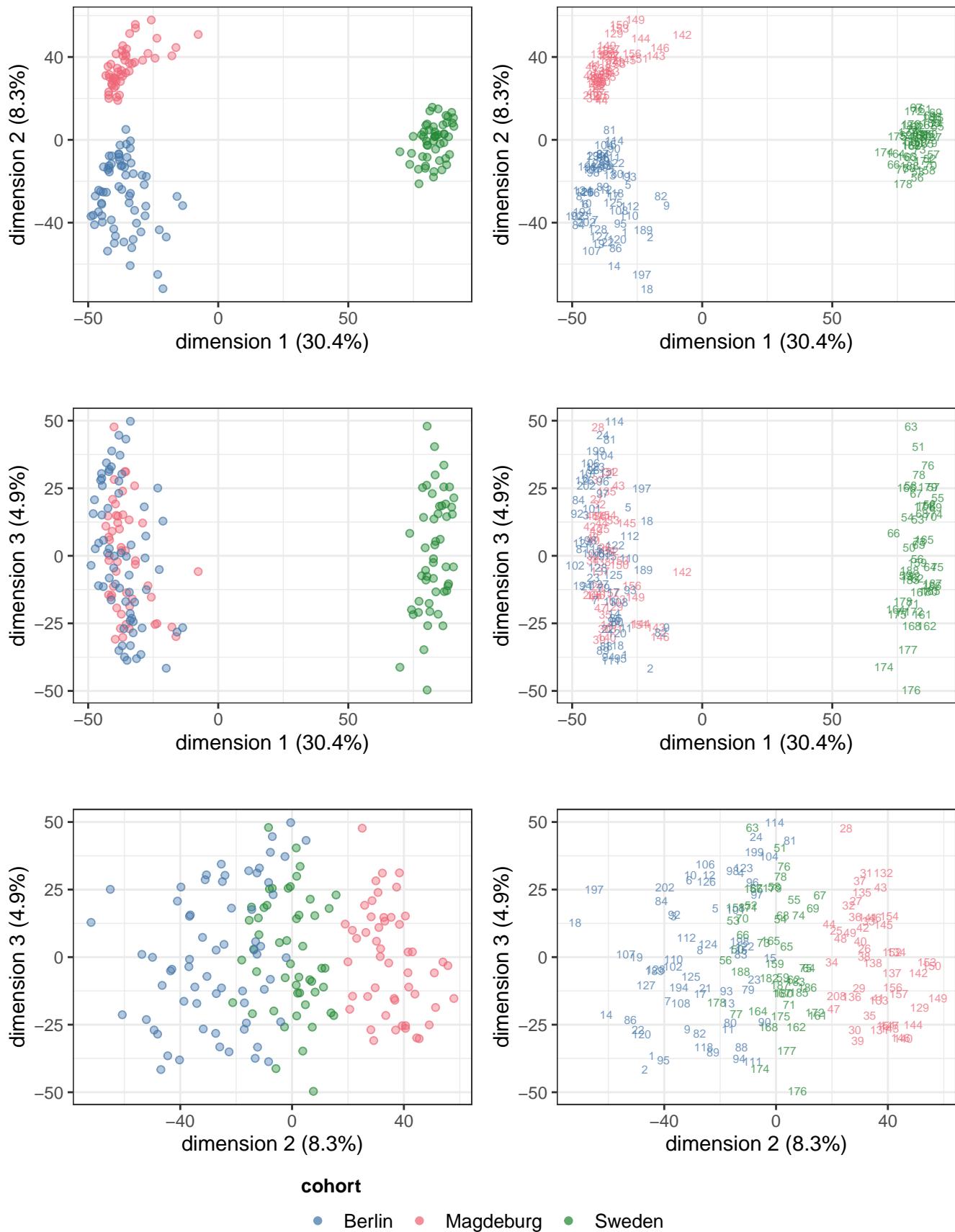
samples are labeled by "sample\_index" (described in samples.xlsx included with MS-DAP output)



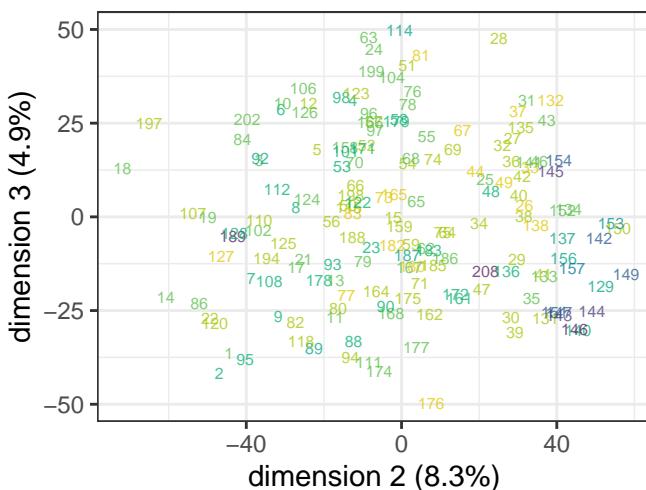
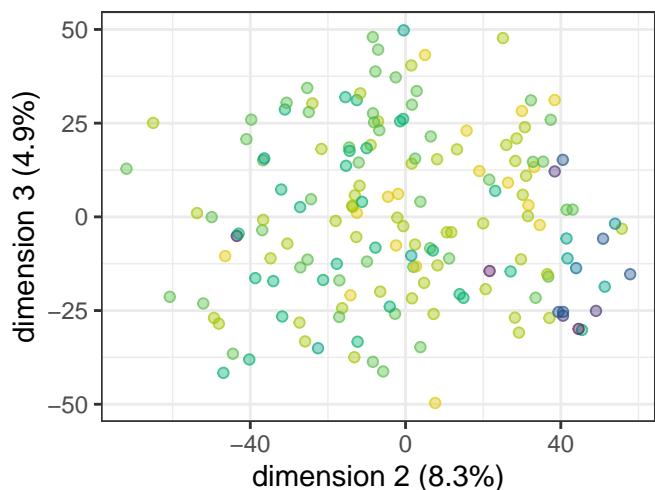
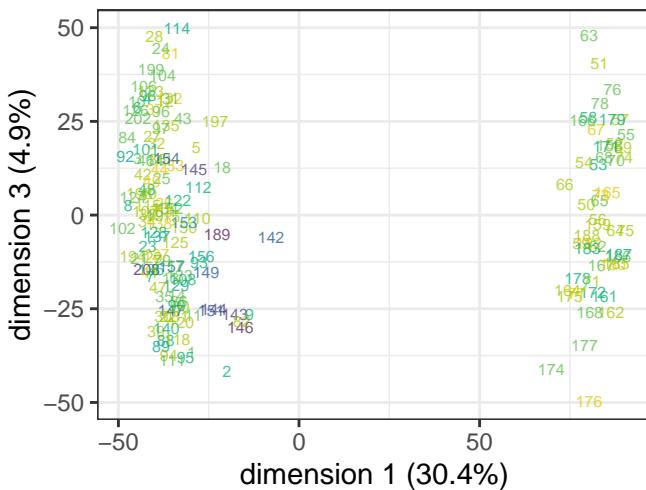
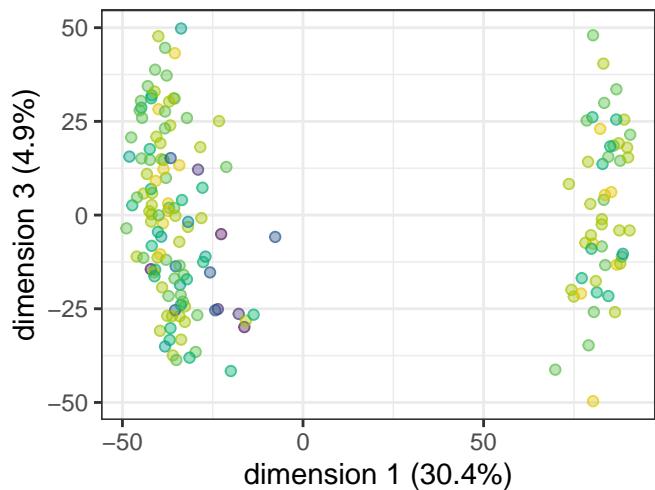
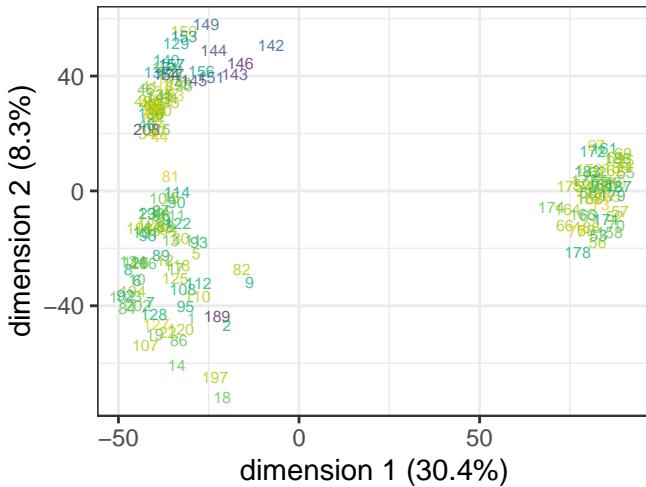
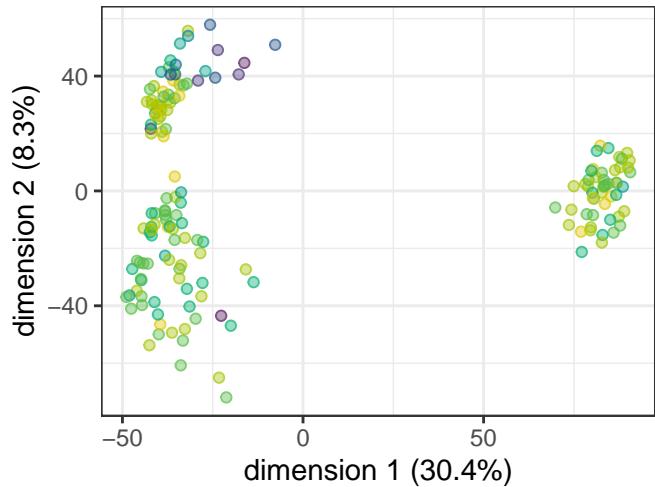
samples are labeled by "sample\_index" (described in samples.xlsx included with MS-DAP output)



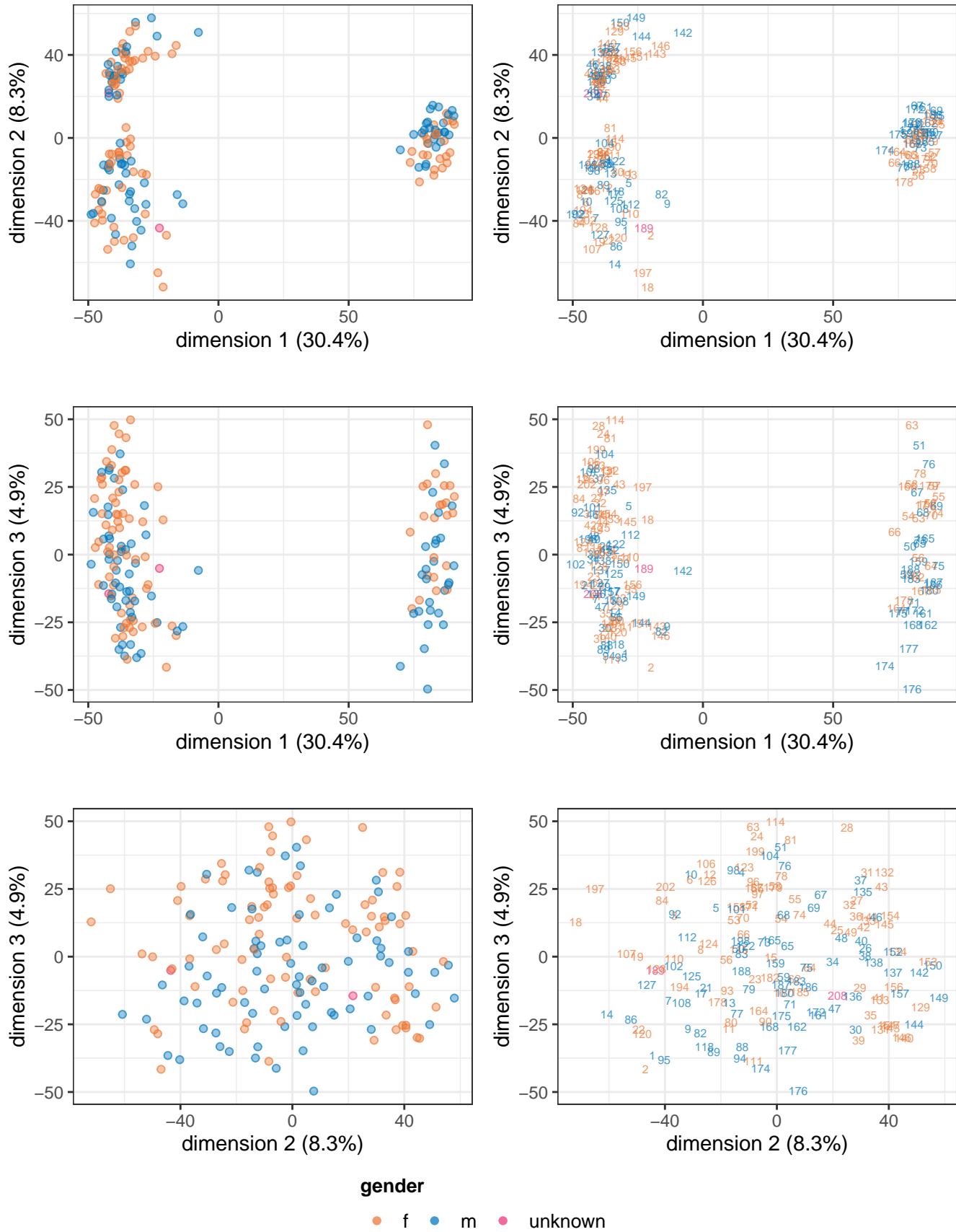
samples are labeled by "sample\_index" (described in samples.xlsx included with MS-DAP output)



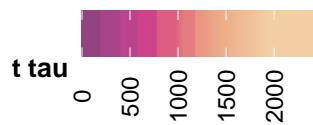
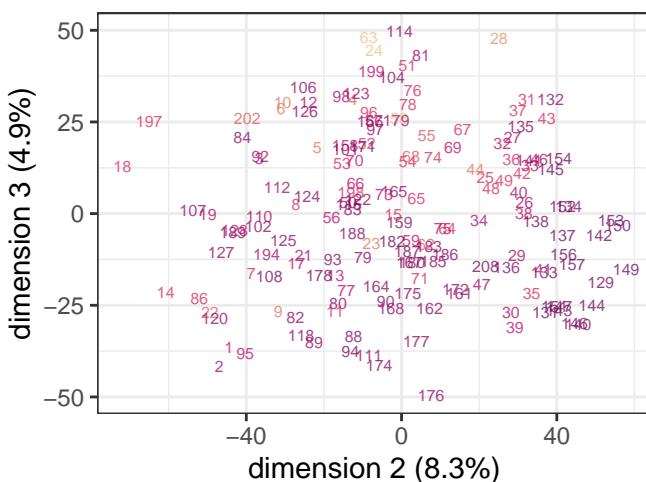
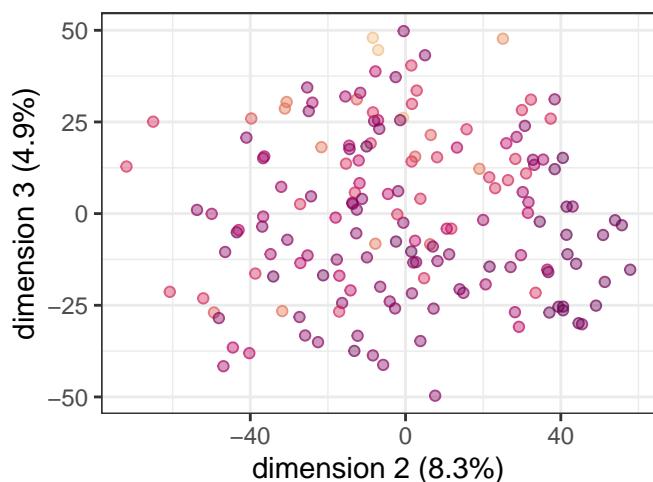
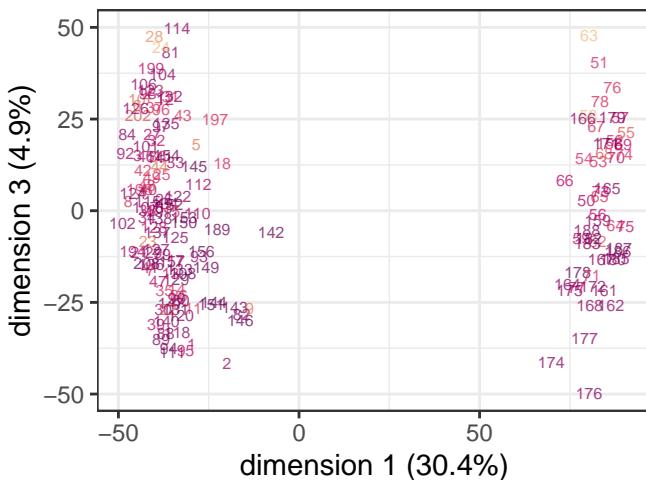
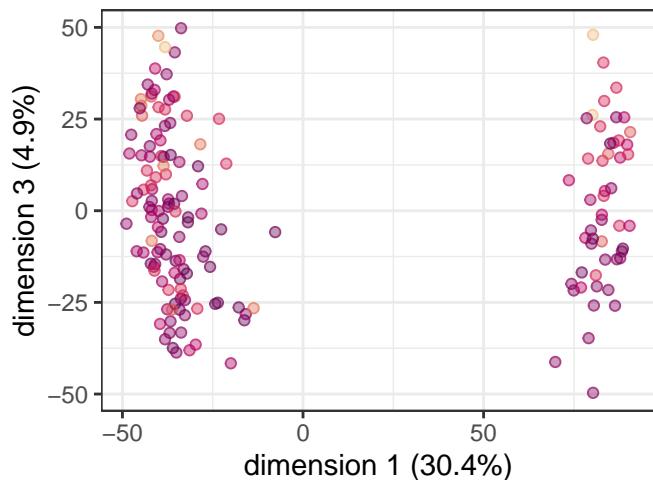
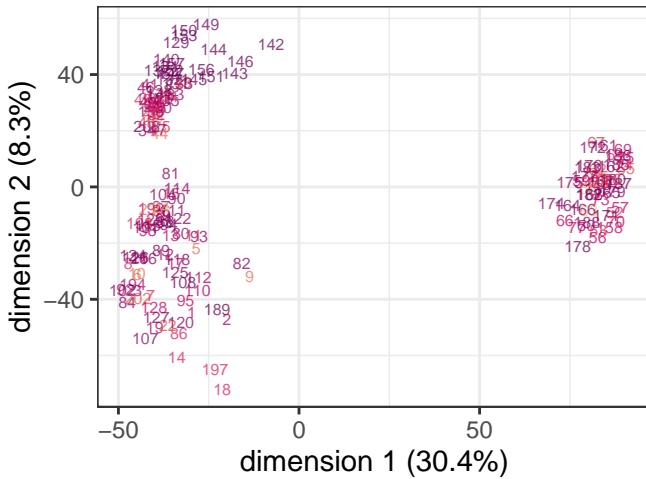
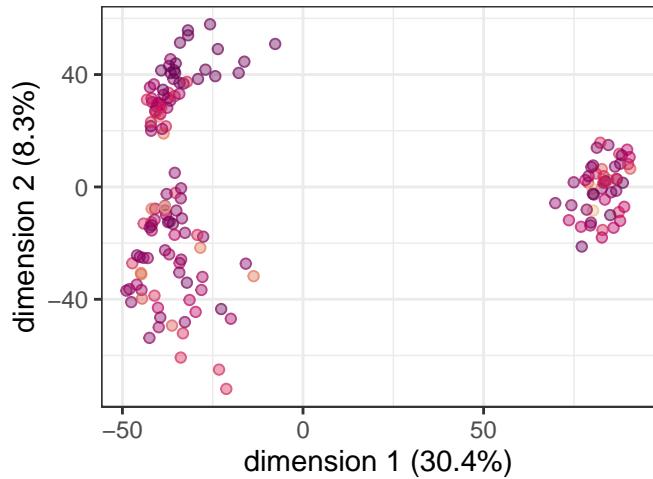
samples are labeled by "sample\_index" (described in samples.xlsx included with MS-DAP output)



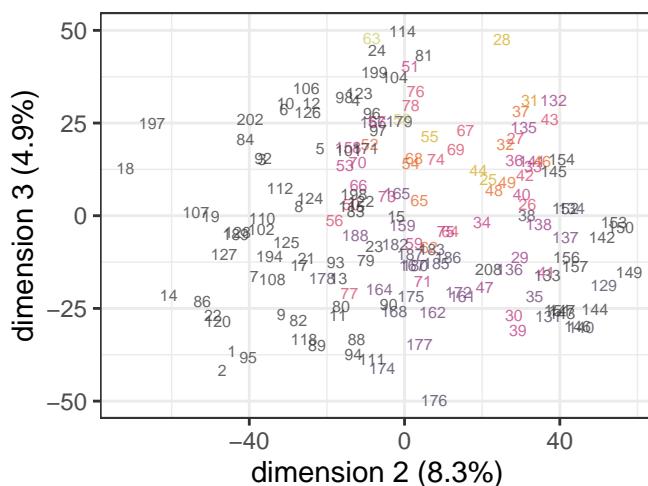
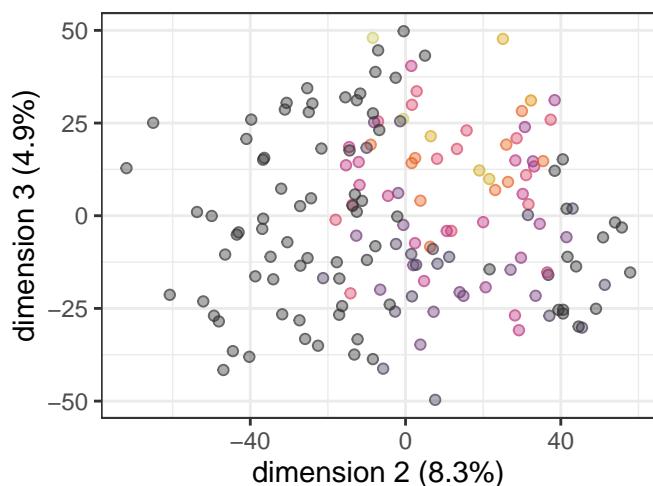
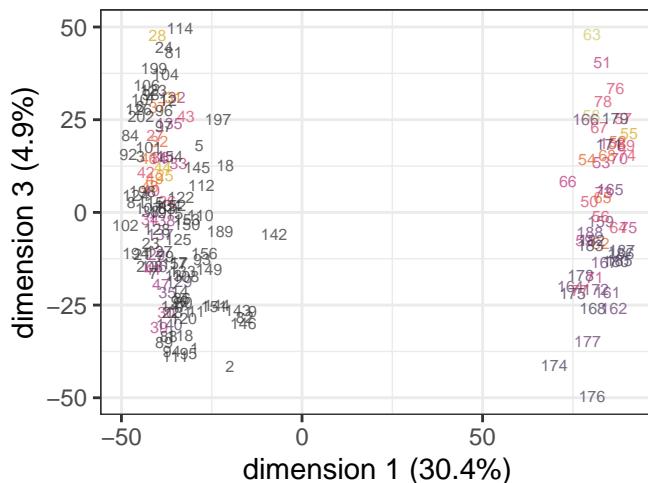
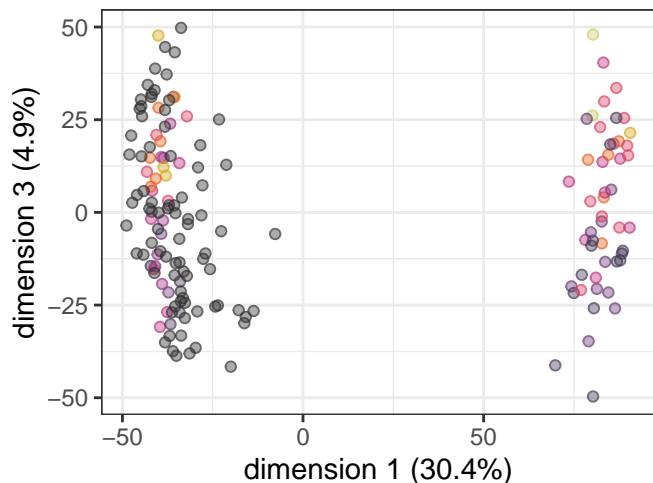
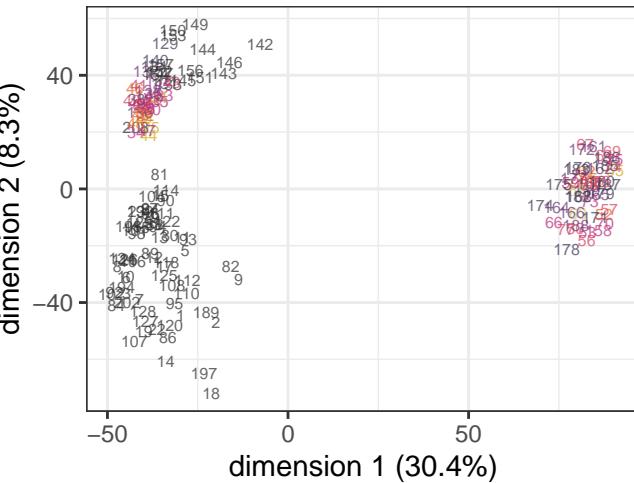
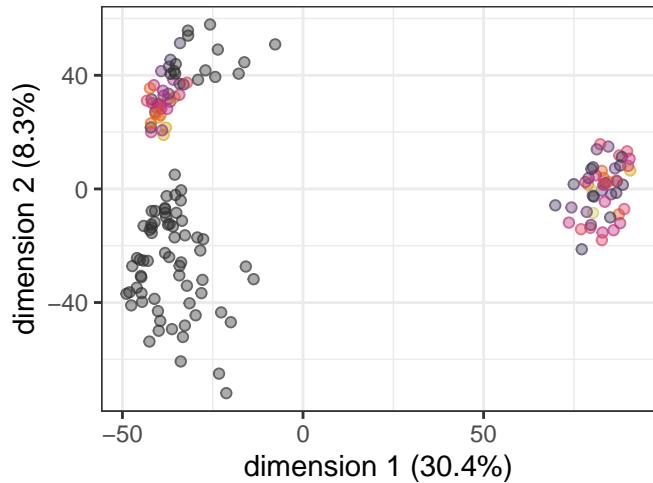
samples are labeled by "sample\_index" (described in samples.xlsx included with MS-DAP output)



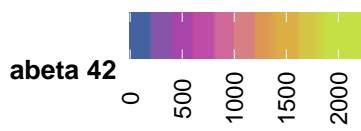
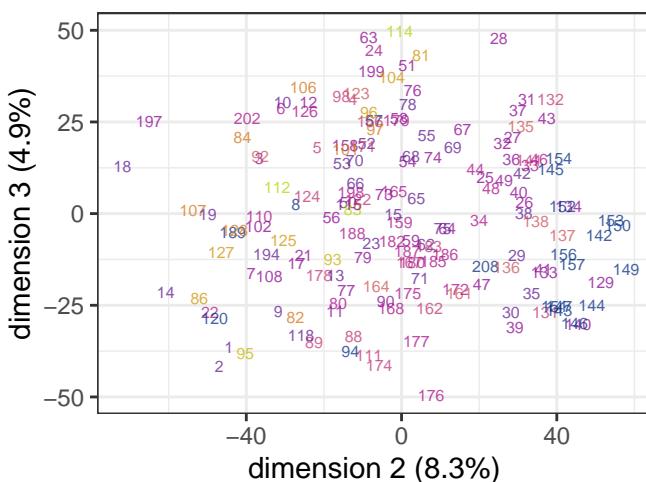
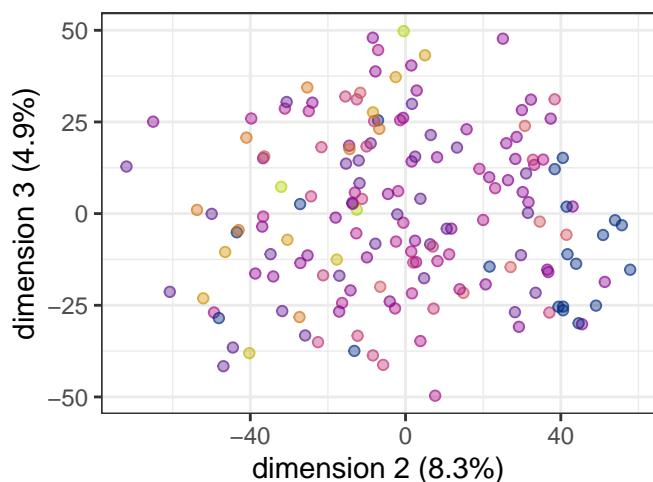
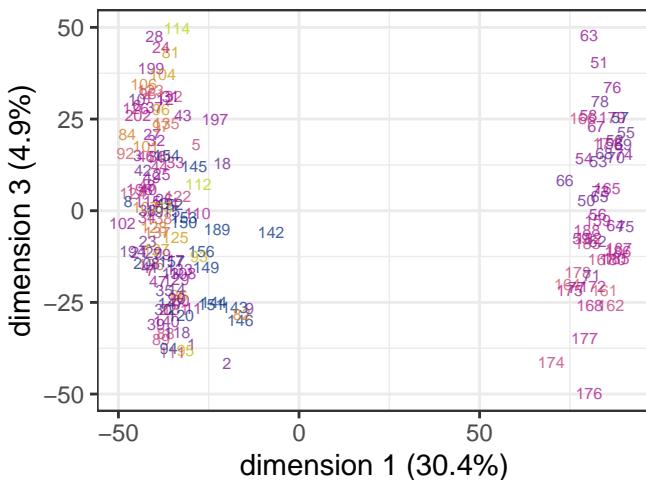
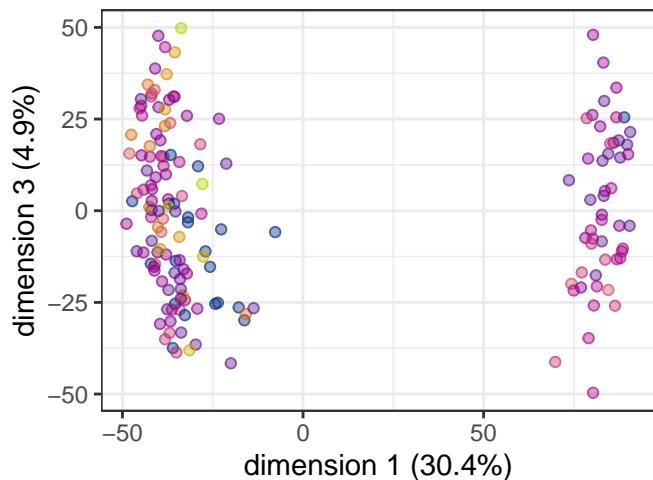
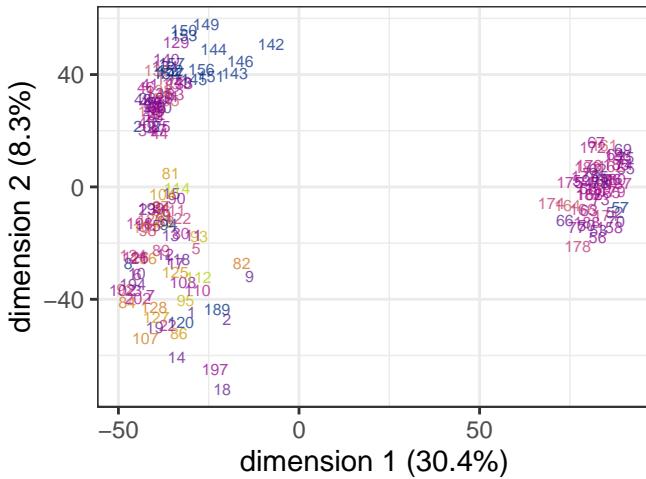
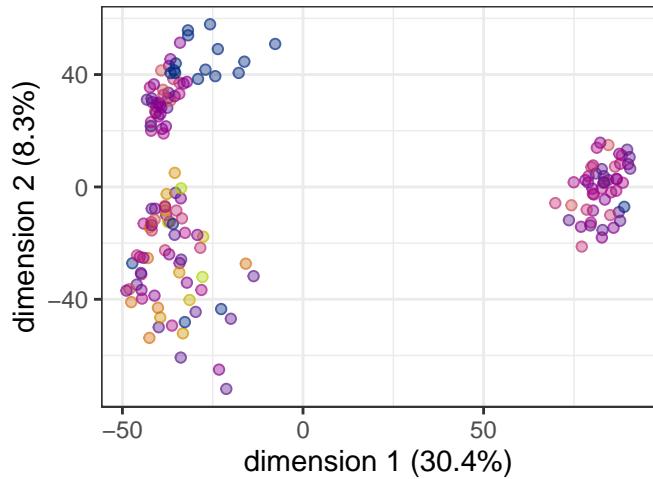
samples are labeled by "sample\_index" (described in samples.xlsx included with MS-DAP output)



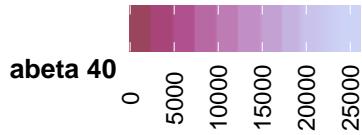
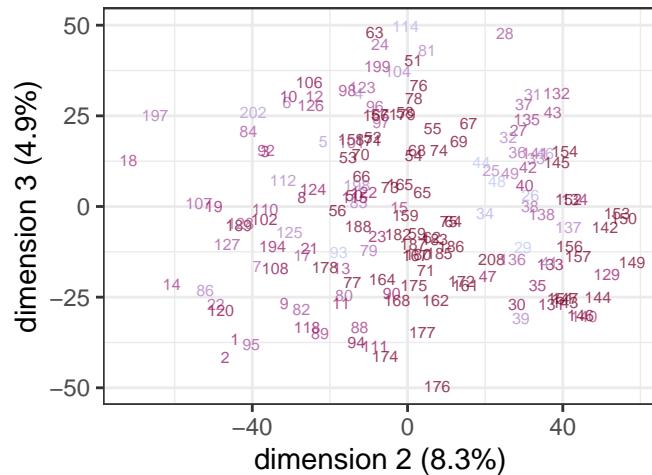
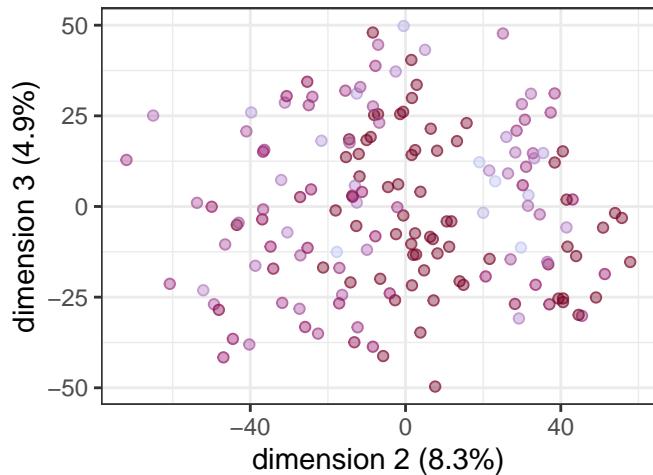
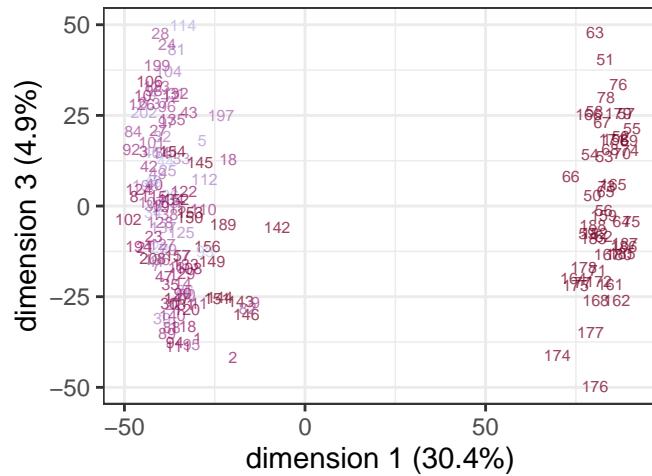
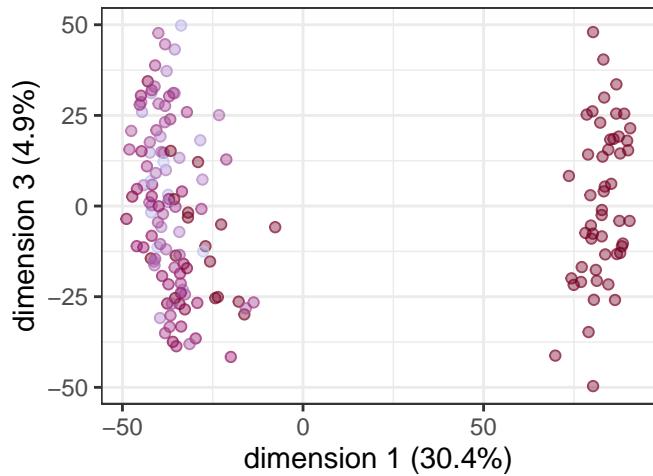
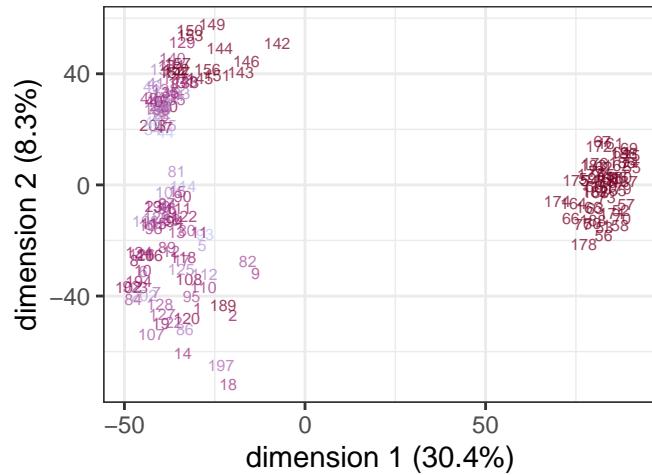
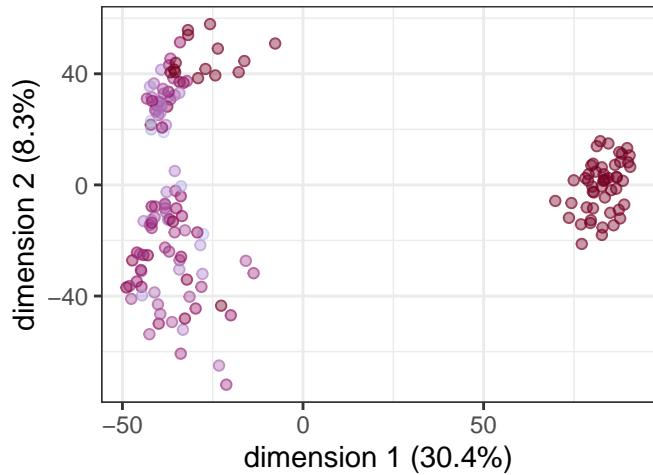
samples are labeled by "sample\_index" (described in samples.xlsx included with MS-DAP output)



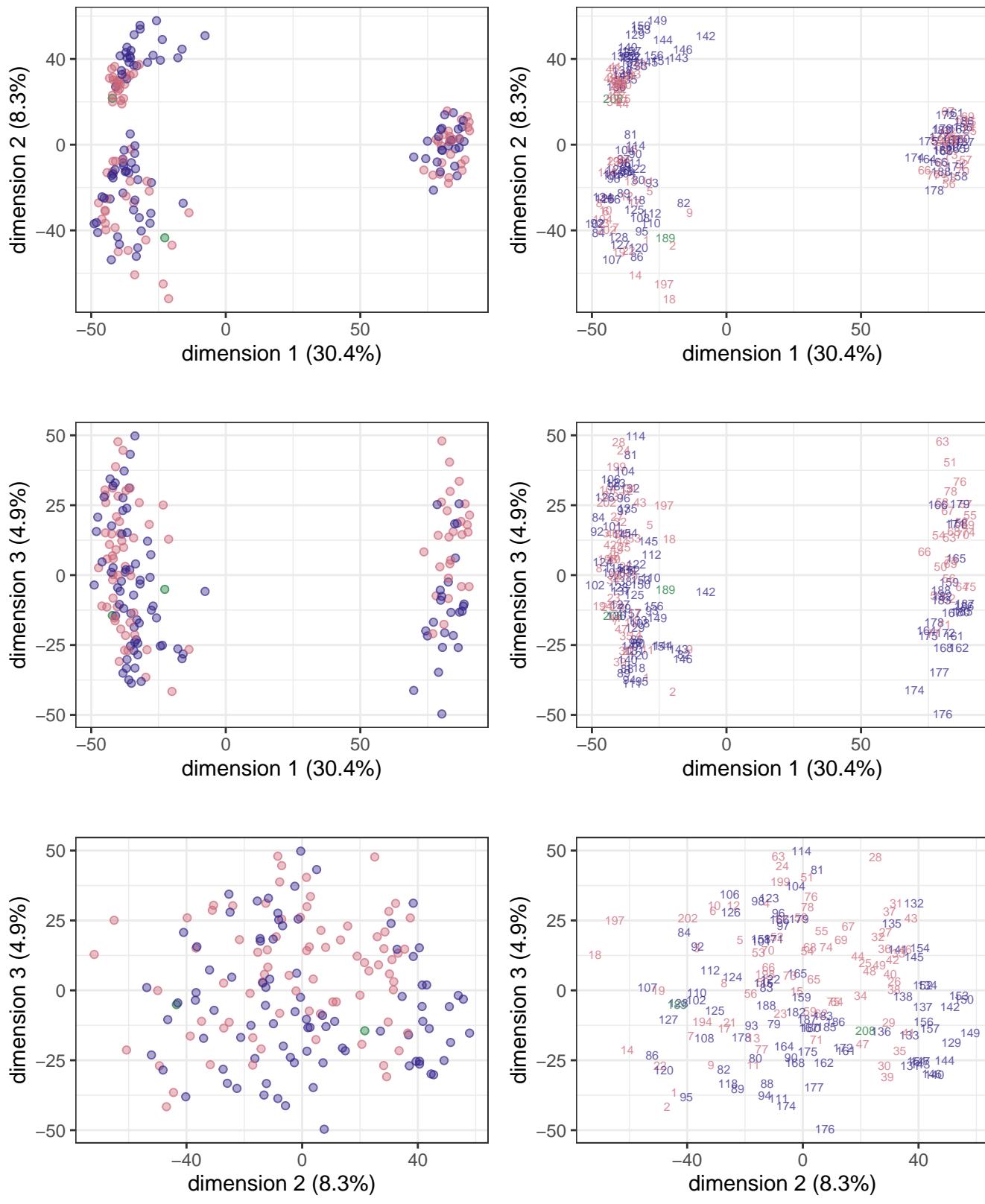
samples are labeled by "sample\_index" (described in samples.xlsx included with MS-DAP output)



samples are labeled by "sample\_index" (described in samples.xlsx included with MS-DAP output)



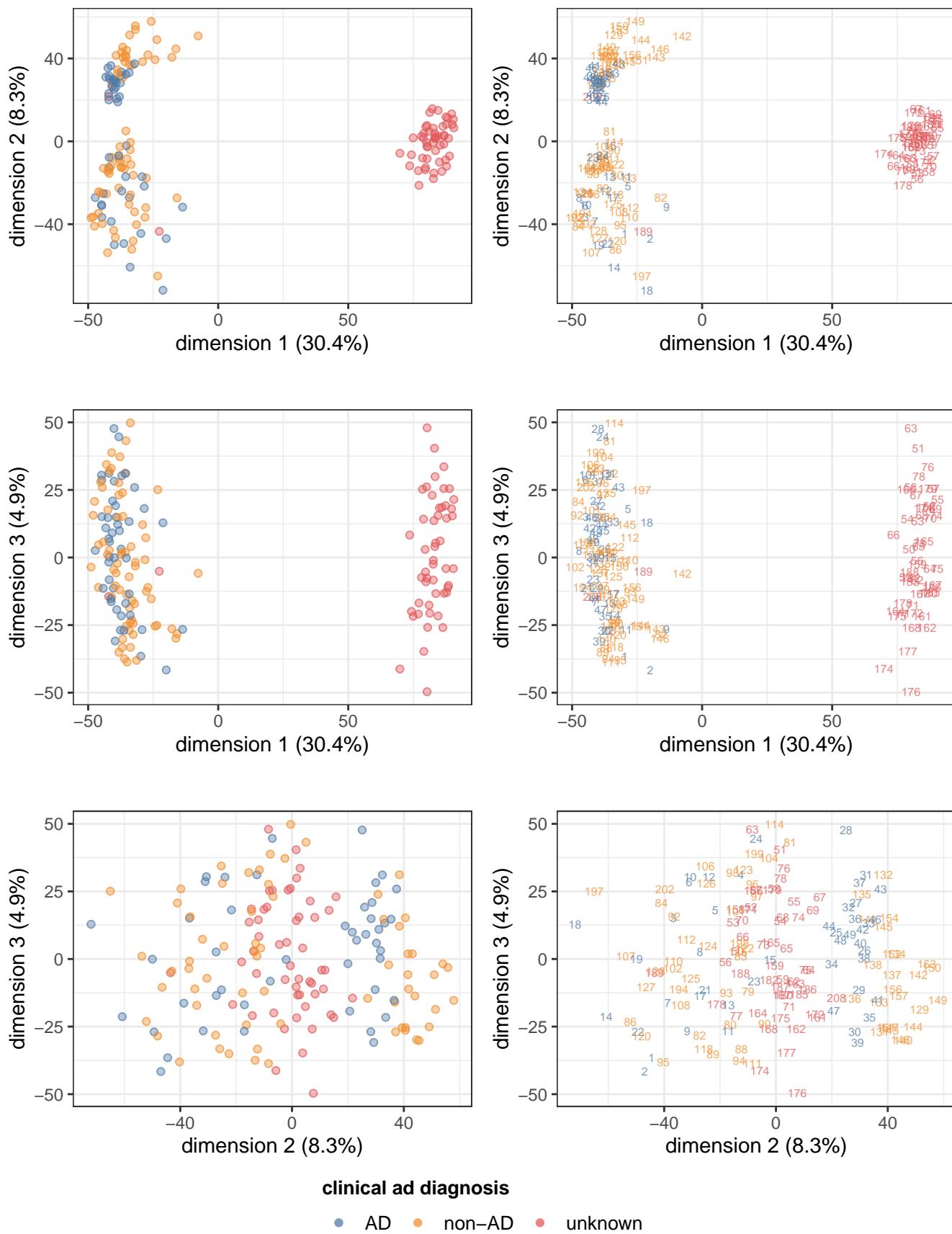
samples are labeled by "sample\_index" (described in samples.xlsx included with MS-DAP output)



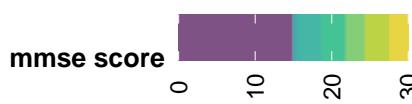
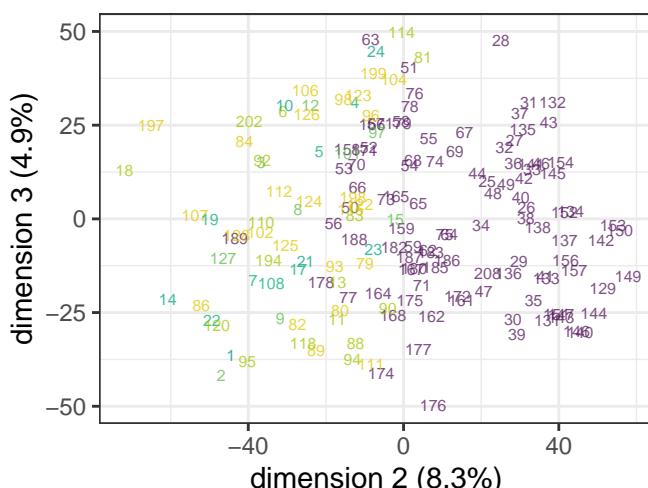
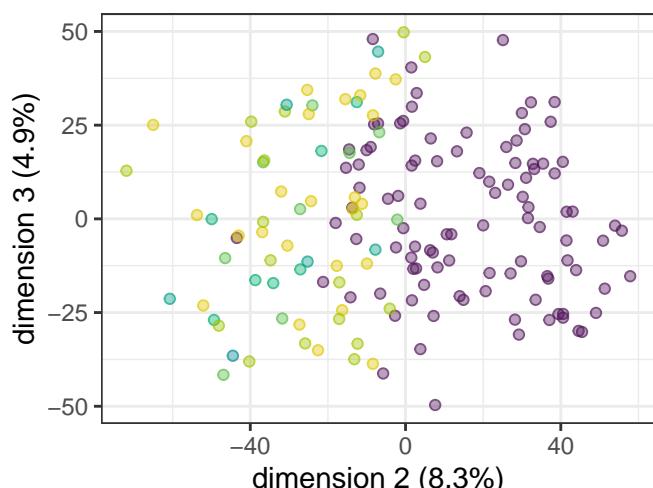
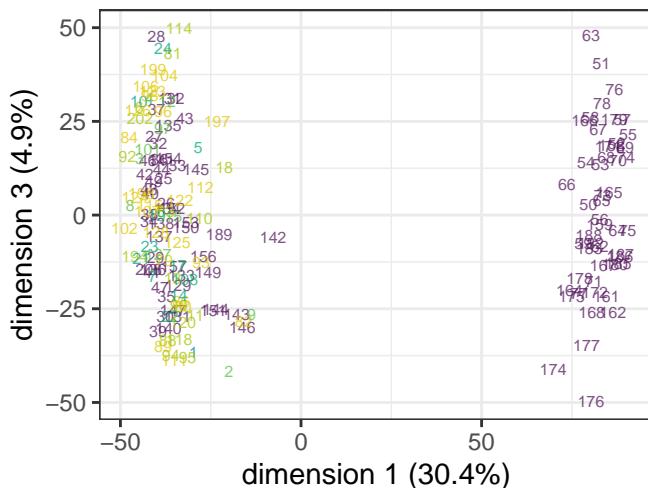
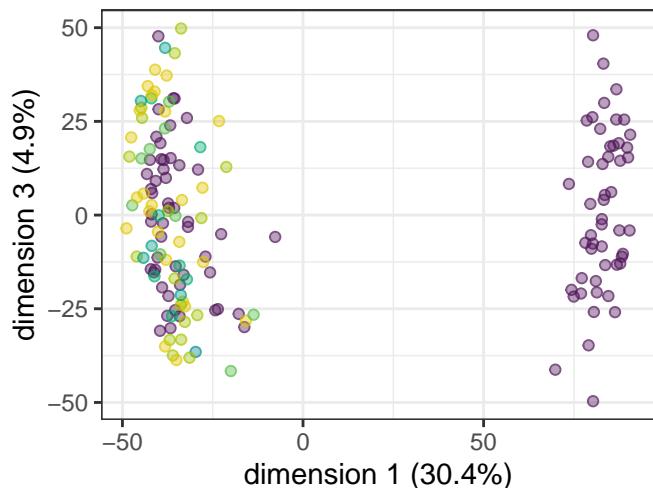
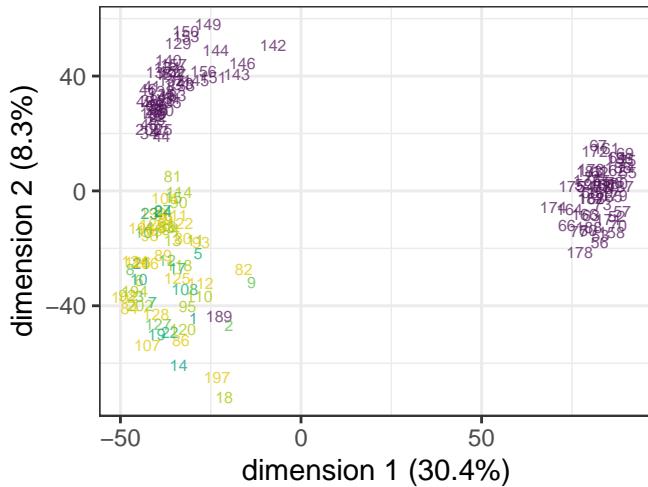
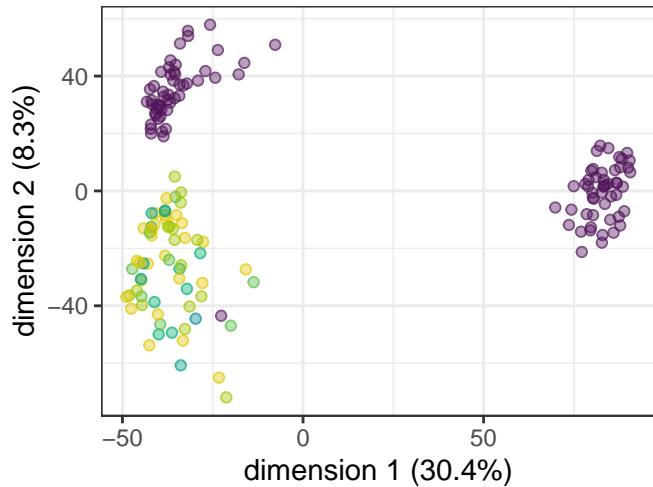
## **biochemical ad classification**

- biochemical AD
- biochemical control
- unknown

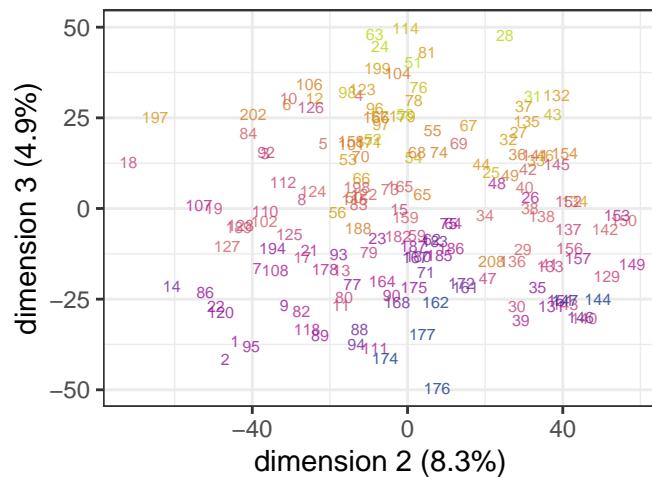
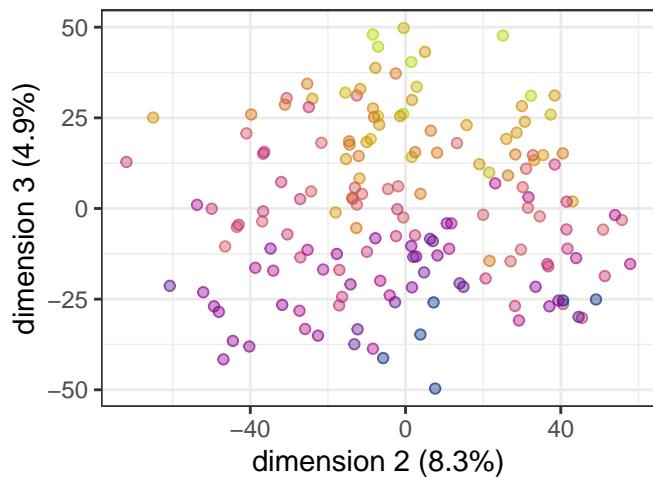
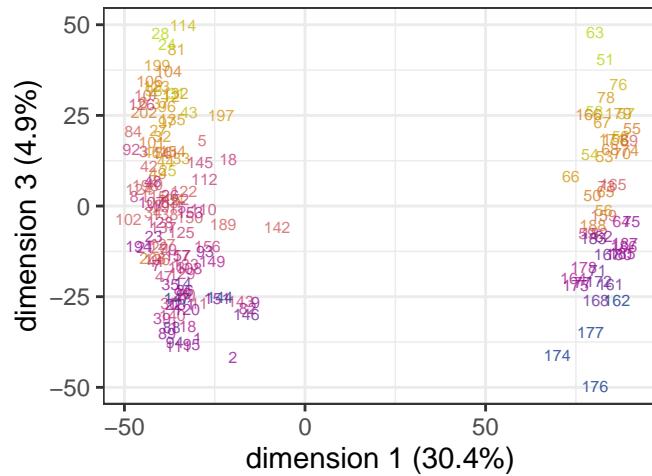
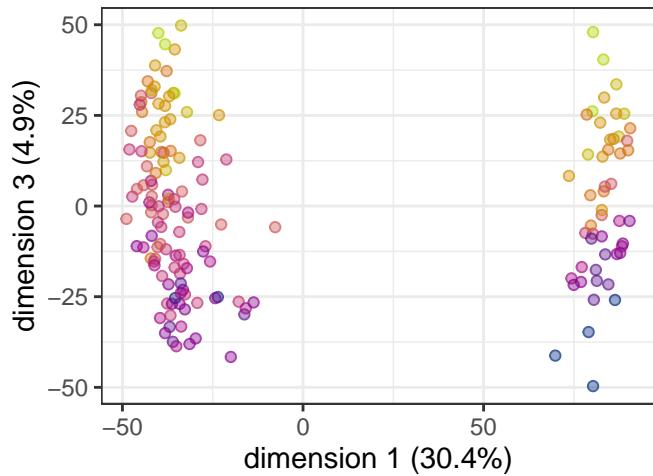
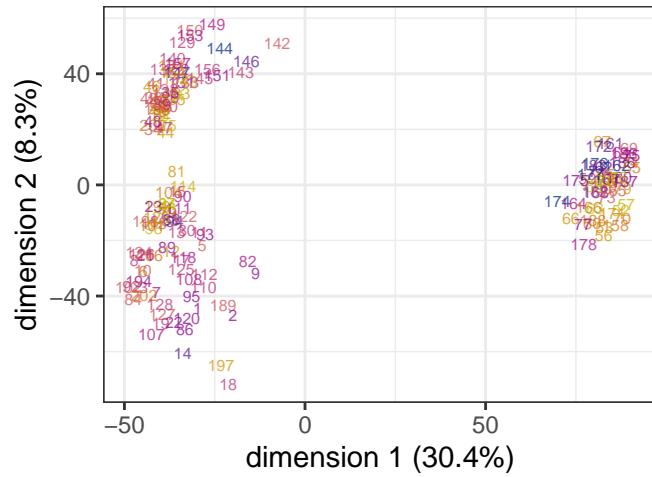
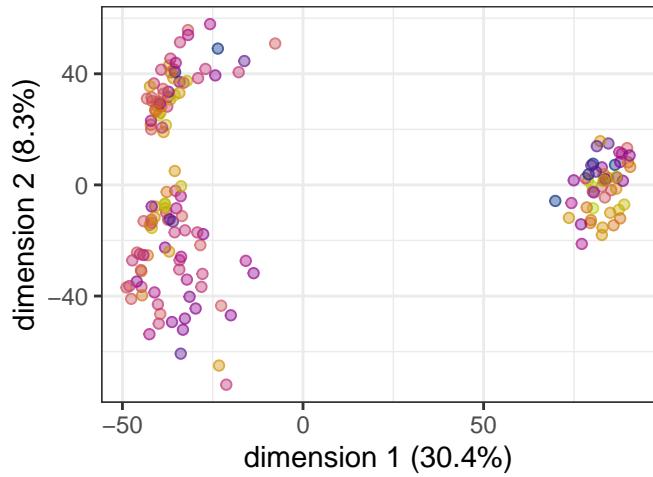
samples are labeled by "sample\_index" (described in samples.xlsx included with MS-DAP output)



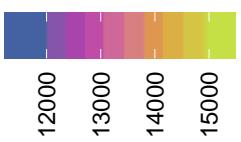
samples are labeled by "sample\_index" (described in samples.xlsx included with MS-DAP output)



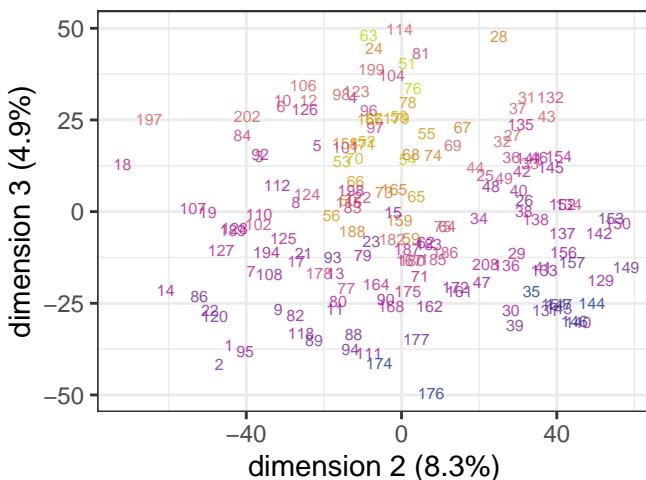
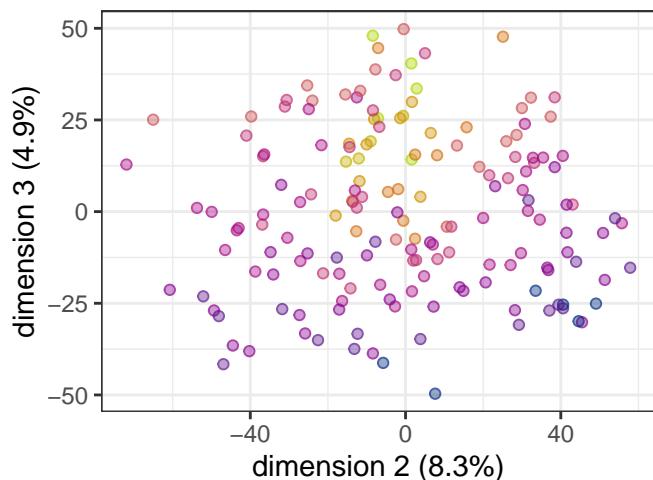
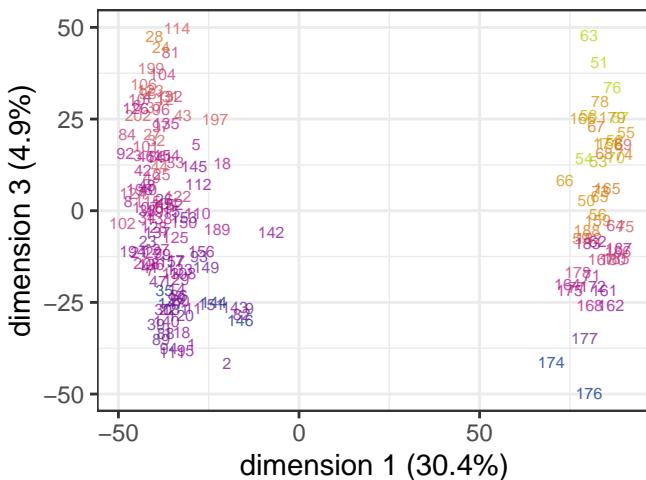
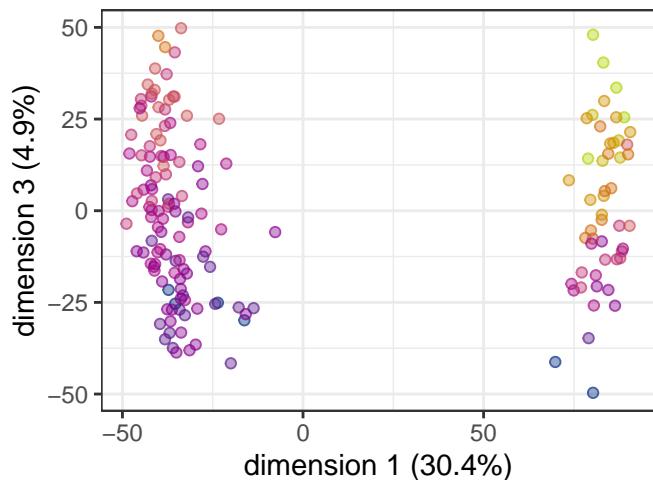
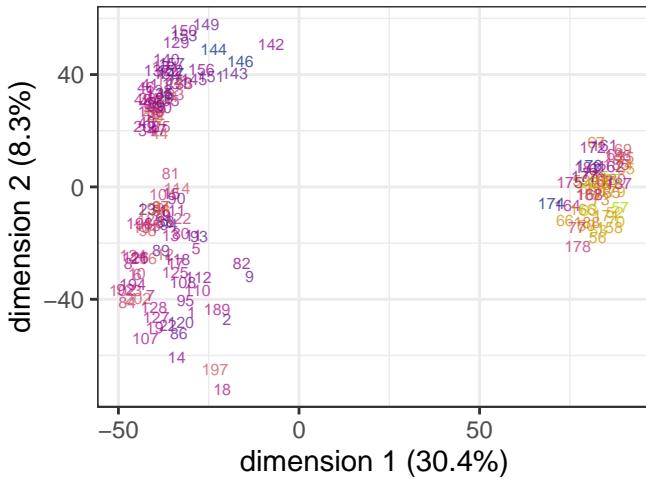
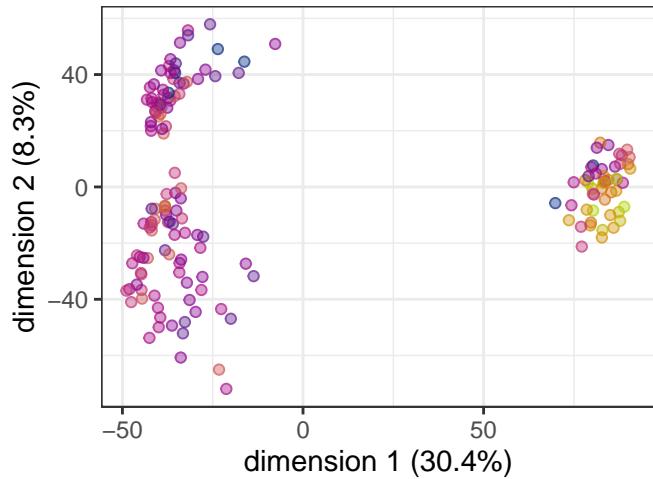
samples are labeled by "sample\_index" (described in samples.xlsx included with MS-DAP output)



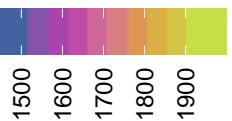
**detected peptides**



samples are labeled by "sample\_index" (described in samples.xlsx included with MS-DAP output)



**detected proteins**



## 2 Differential abundance analysis

### goal: maximize reliable features for quantification

In a pairwise analysis of two groups of samples, only peptides with N data-points in both groups are used for quantitative analysis (where N = defined by user settings). For example; if peptide  $p$  is consistently quantified in sample groups A and B but not in C/D/E, it can be used when comparing group A *versus* group B but should not be used in any other group comparisons. This approach is particularly suited to maximize the number of peptides used for statistical analysis in experimental designs with many sample groups.

A common alternative strategy is a global filtering approach where peptides are selected based on their properties in the overall dataset (eg; present in x% of samples or x% of replicates in all groups) and subsequently the resulting data matrix is used for all downstream statistical analyses. In the example above where peptide  $p$  is present in a subset of sample groups,  $p$  would either be left out (not present in majority of samples in entire dataset) or erroneously used when applying t-statistics to groups B and C (since  $p$  is not present in group C, it may differentially detected but there are no features available for quantitative analysis)

### 2.1 control\_Berlin vs AD\_Berlin

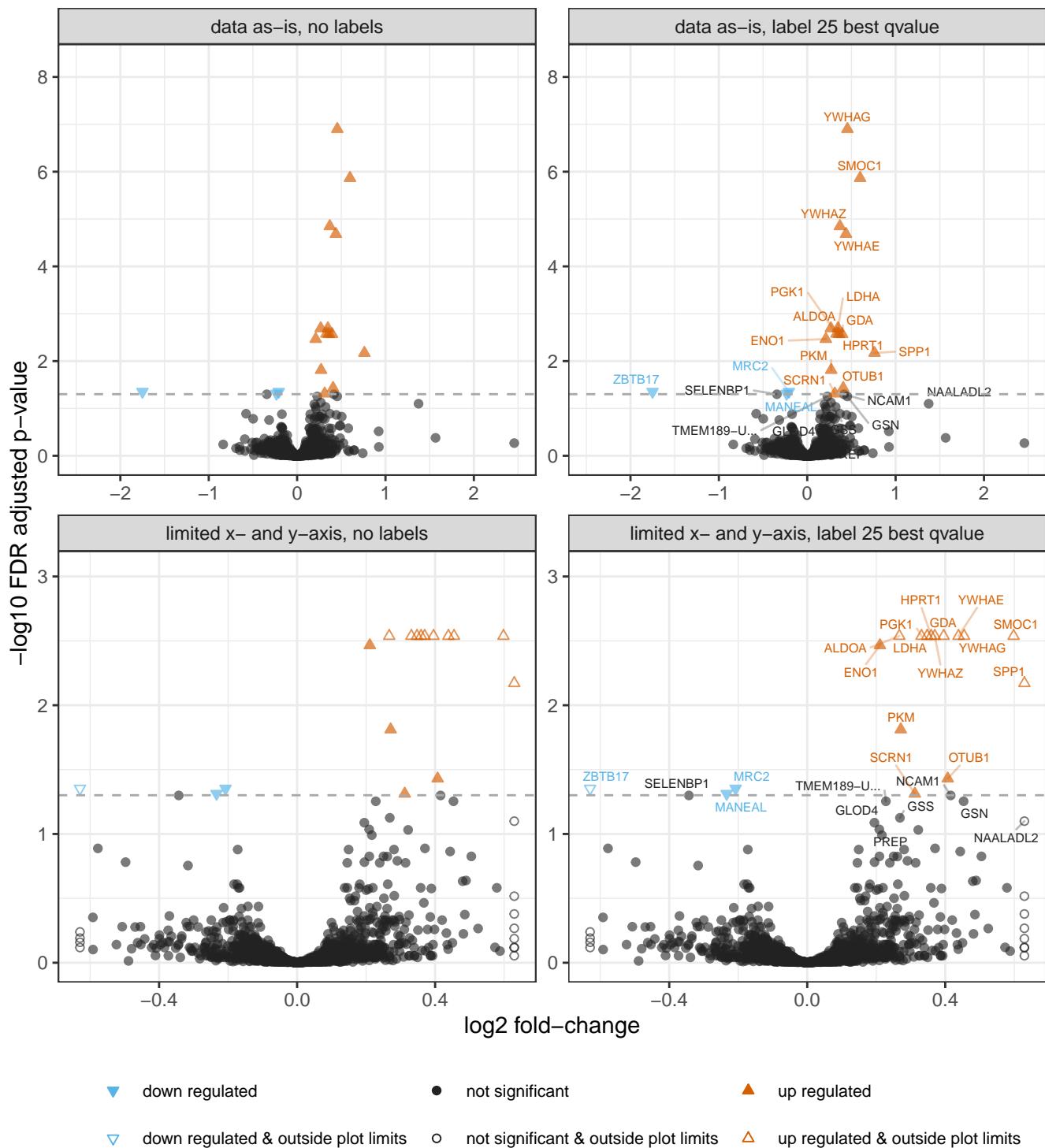
- **user setting:** using ‘filter by contrast’ peptide filtering approach
- 14318 peptides in 1662 proteins remain in the current contrast after peptide filters and are used for the statistical analysis in this section
- qvalue threshold: 0.05
- log2 foldchange threshold: 0

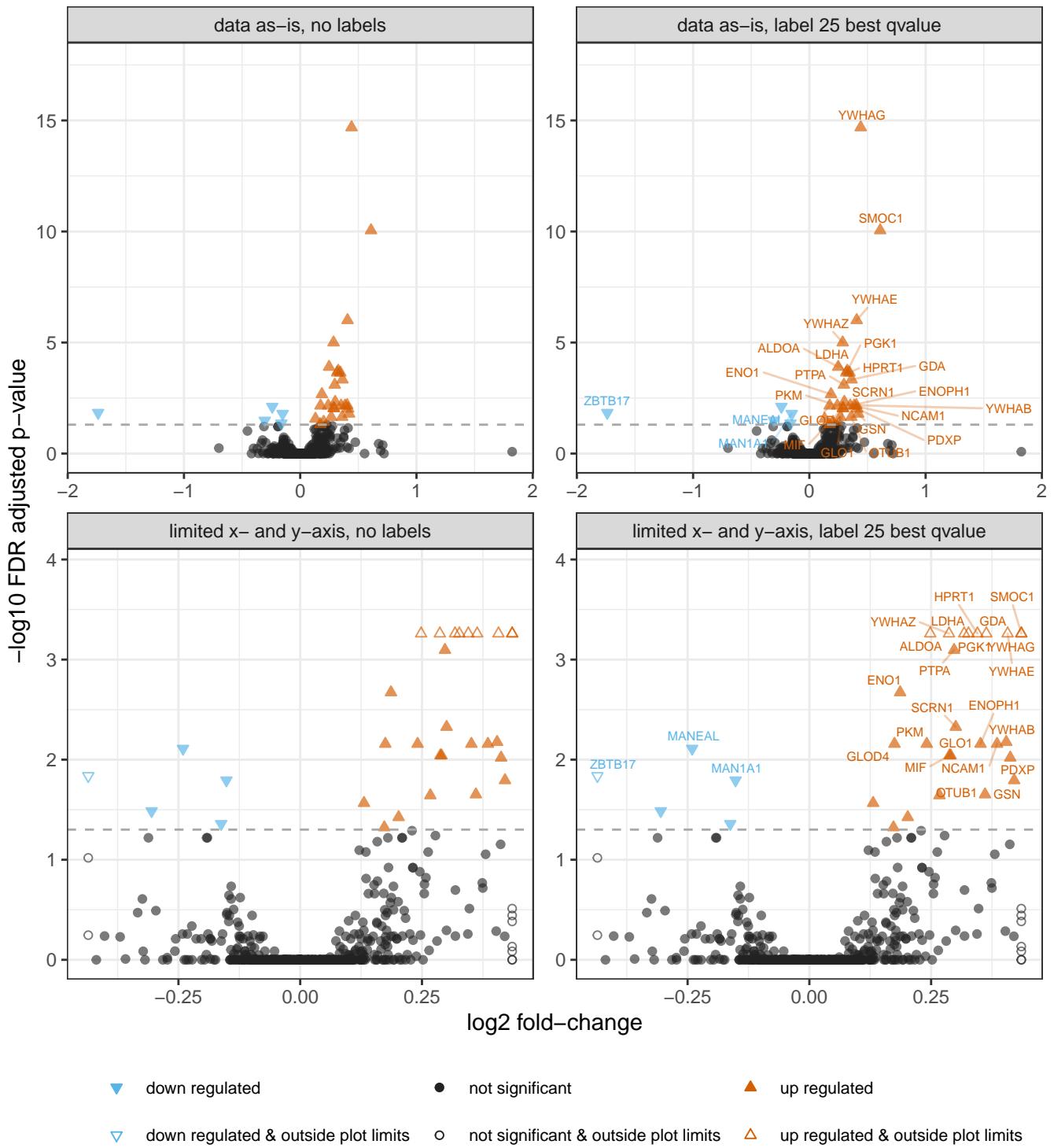
#### 2.1.1 volcano

The plot title shows the statistical model and contrast (sample groups in the comparison). Left- and right-side figure panels on each row represent the same figure without and with labels for the 25 proteins with lowest p-value.

Bottom figure panels have limited x- and y-axis. For datasets with a small number of strong outliers in p-value or fold-change, which may have a profound effect on the plot scales, this allows inspection of the remainder of the volcano plot without disproportionate influence by ‘extreme’ values.

Labels for proteins that are more than 12 characters long are truncated for visual clarity (indicated by trailing ...). For protein identifiers that are ambiguous, e.g. a protein-group with assigned genes “gene1a;gene1b”, only the first label/ID is shown for visual clarity (indicated by trailing \*).

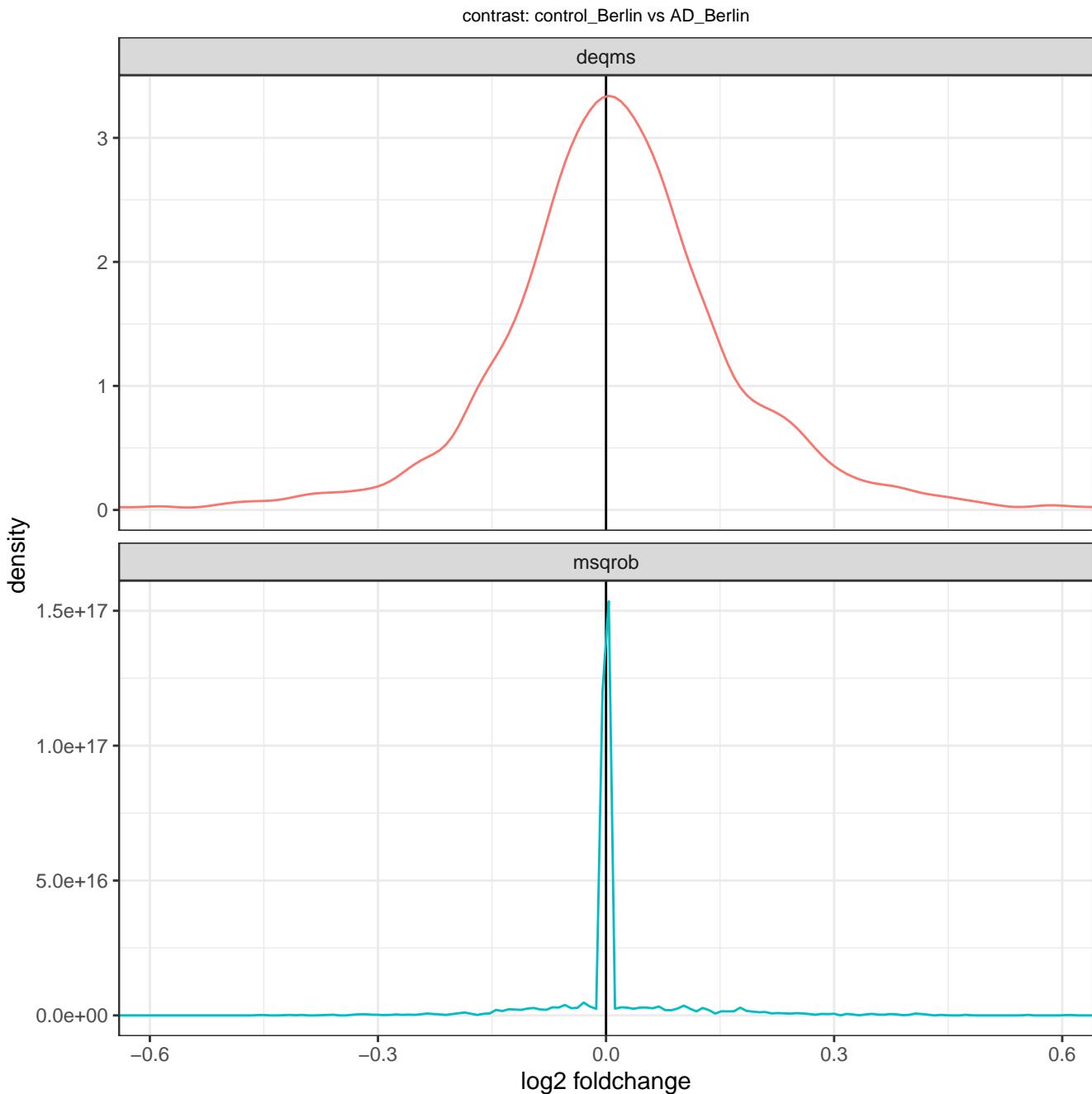




### 2.1.2 foldchange distribution

Distributions of estimated foldchanges produced by the statistical models. If the mode is far from 0, consider alternative normalization strategies. Do note the scale on the x-axis, for some experiments the foldchanges are very low which in turn may exaggerate this figure.

*note; the MSqRob model tends to assign zero (log)foldchange for proteins with minor difference between conditions where the model is very sure the null hypothesis cannot be rejected (shrinkage by the ridge regression model). As a result, many foldchanges will be zero and the density plot for MSqRob may look like a spike instead of the expected Gaussian shape observed in other models*



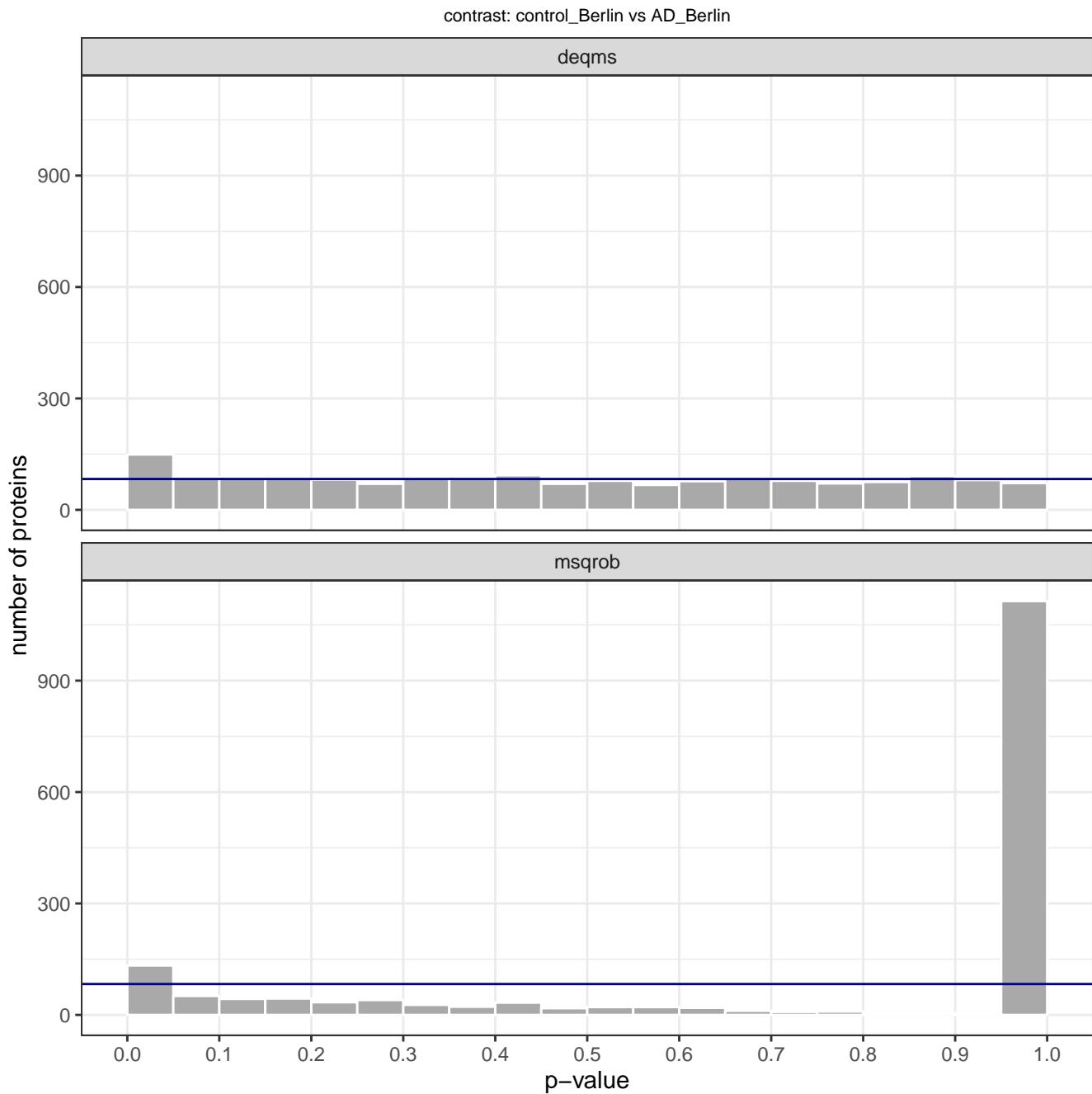
### 2.1.3 p-value distribution

Histogram of p-values computed by differential expression analysis algorithms, as-is, for quality-control inspection. The horizontal line indicates the expected counts assuming a uniform distribution (total number of p-values divided by number of histogram bins)

See further: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6164648/>

See further: <http://varianceexplained.org/statistics/interpreting-pvalue-histogram/>

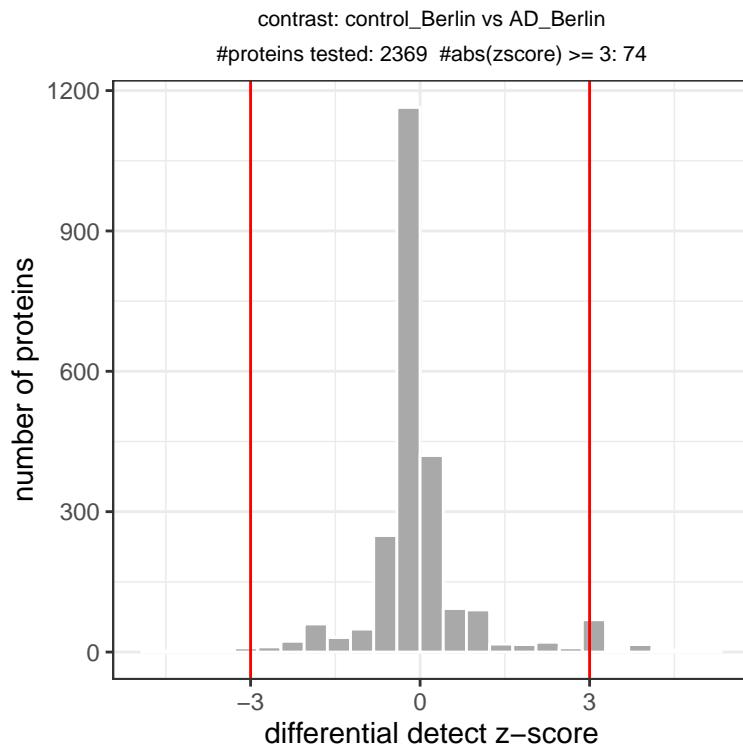
*note; the MSqRob and MS-Empire models often yield p-value distributions that show a large peak at p-value 1, these are typically proteins with estimated log foldchanges at/near zero where these models are very sure the null hypothesis cannot be rejected*



#### 2.1.4 differential detect

Some proteins may not have peptides with sufficient data points over samples to be used for differential expression analysis (depending on the user-defined filtering criteria in how many replicates peptides should be observed), but do show a strong difference in the number of detected peptides between sample groups. In some proteomics experimental designs, for example a wildtype-knockout APMS study, those are interesting proteins. The DEA based on peptide abundance values (volcano plots above) are the main result for differential testing in MS-DAP but as a situationally useful tool MS-DAP also includes a ‘protein detection’ z-score, based on the number of times a peptide for each protein was detected per sample group (/experimental condition), as an alternative means of differential testing.

Below figure shows the distribution of these scores with thresholds at 3 std. Both the z-scores and the counts these are based upon are available in the statistical result Excel table.



## 2.2 control\_Magdeburg vs AD\_Magdeburg

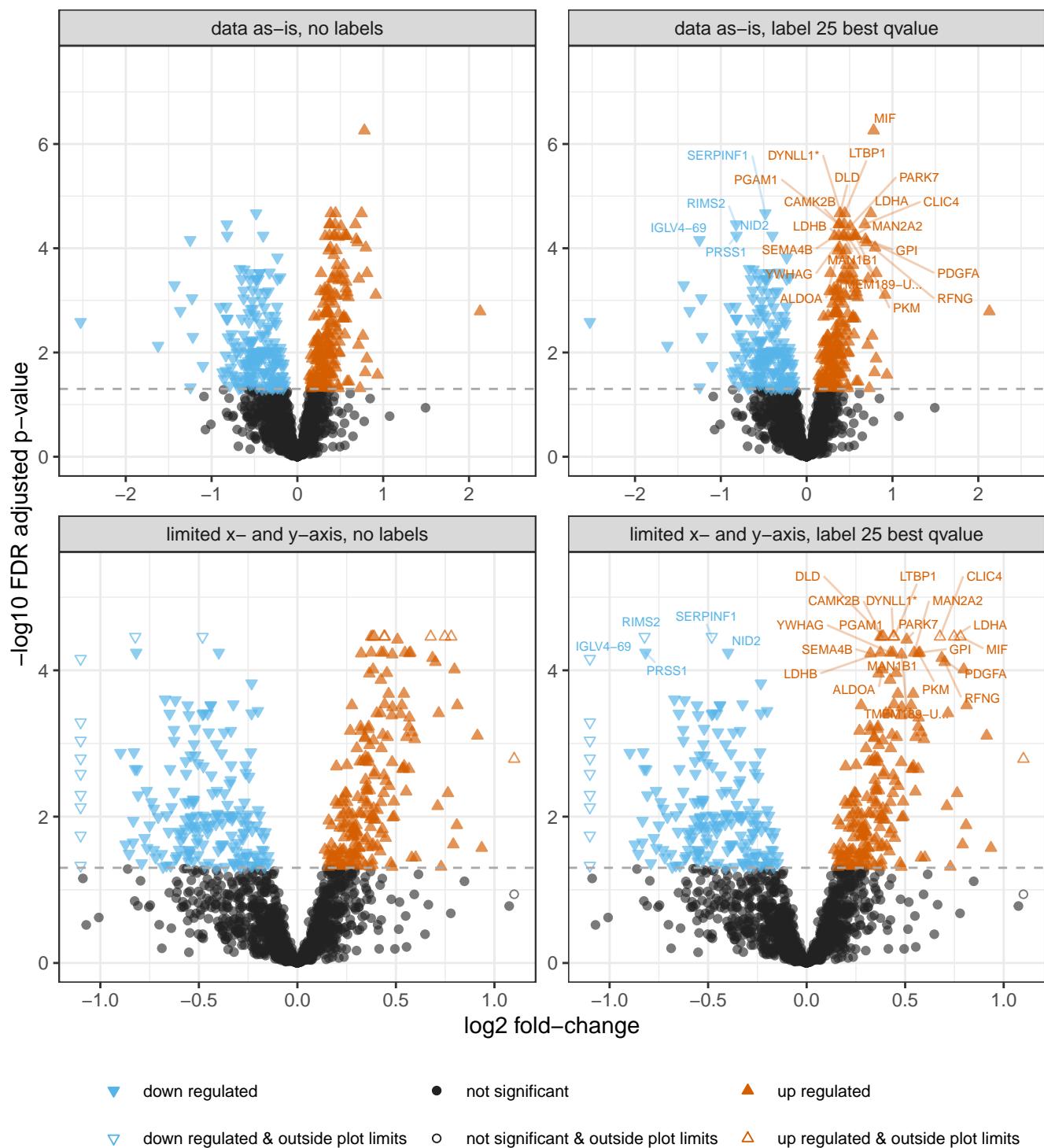
- **user setting:** using ‘filter by contrast’ peptide filtering approach
- 14255 peptides in 1617 proteins remain in the current contrast after peptide filters and are used for the statistical analysis in this section
- qvalue threshold: 0.05
- log2 foldchange threshold: 0

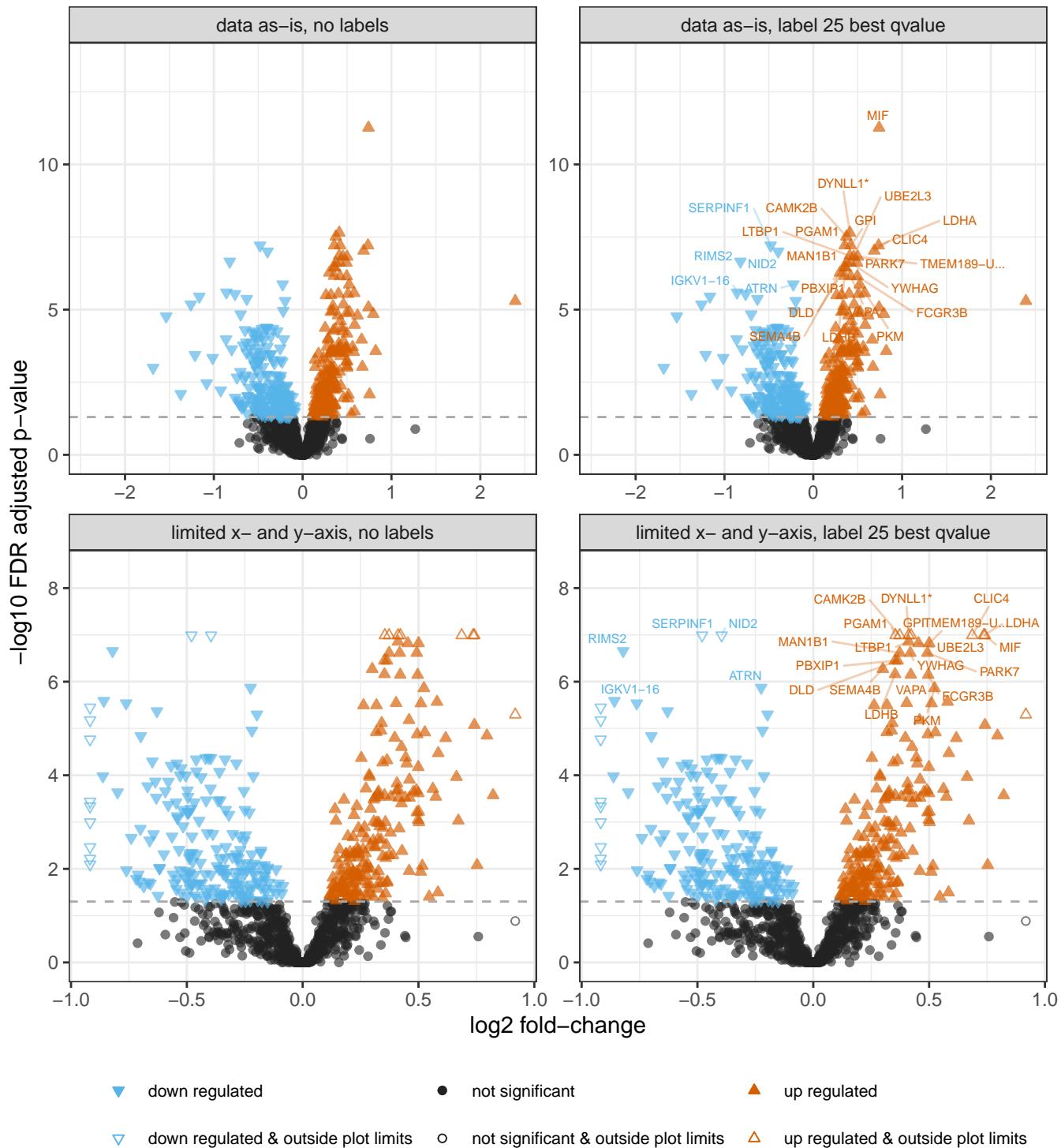
### 2.2.1 volcano

The plot title shows the statistical model and contrast (sample groups in the comparison). Left- and right-side figure panels on each row represent the same figure without and with labels for the 25 proteins with lowest p-value.

Bottom figure panels have limited x- and y-axis. For datasets with a small number of strong outliers in p-value or fold-change, which may have a profound effect on the plot scales, this allows inspection of the remainder of the volcano plot without disproportionate influence by ‘extreme’ values.

Labels for proteins that are more than 12 characters long are truncated for visual clarity (indicated by trailing ...). For protein identifiers that are ambiguous, e.g. a protein-group with assigned genes “gene1a;gene1b”, only the first label/ID is shown for visual clarity (indicated by trailing \*).

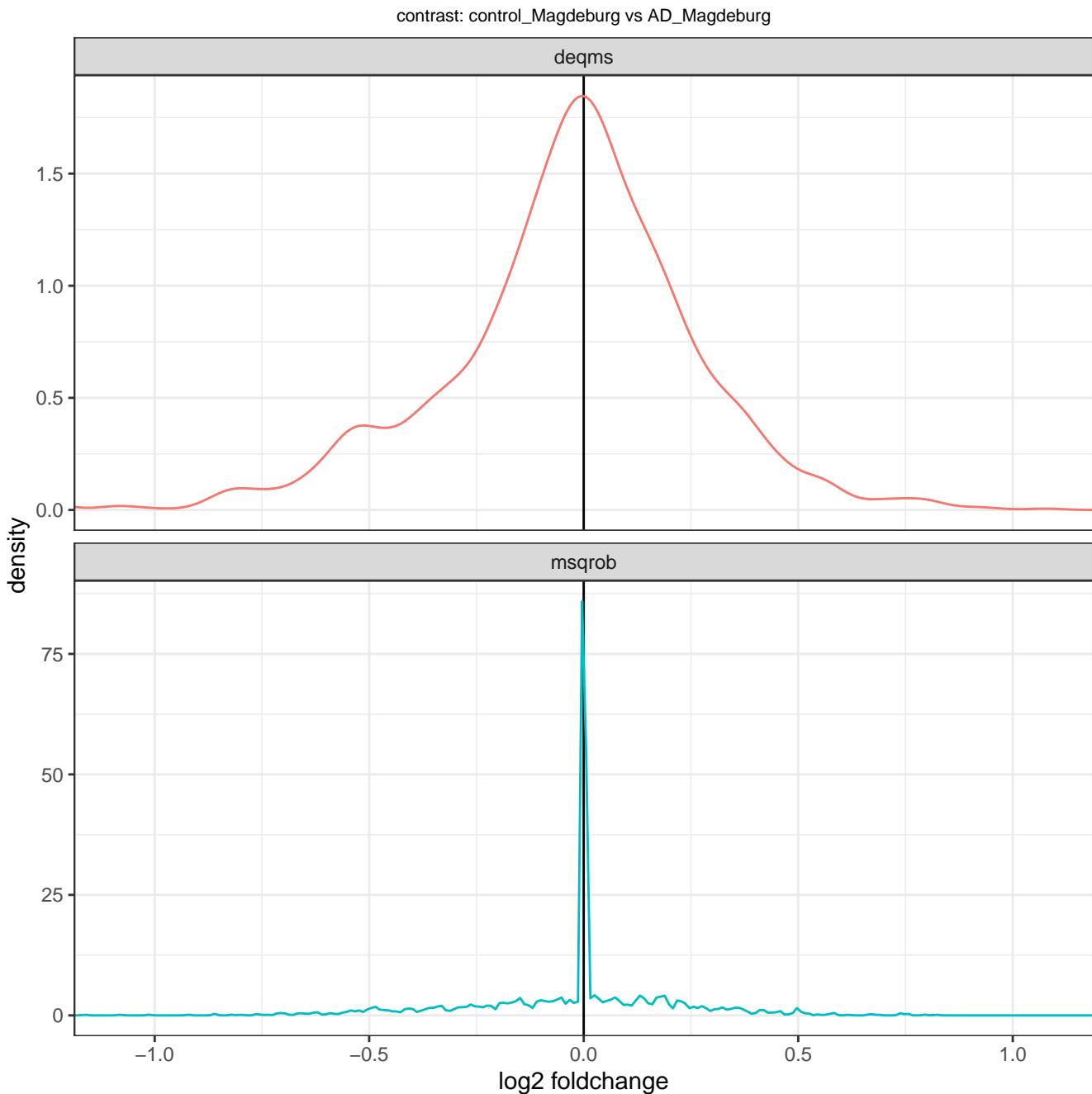




## 2.2.2 foldchange distribution

Distributions of estimated foldchanges produced by the statistical models. If the mode is far from 0, consider alternative normalization strategies. Do note the scale on the x-axis, for some experiments the foldchanges are very low which in turn may exaggerate this figure.

*note; the MSqRob model tends to assign zero (log)foldchange for proteins with minor difference between conditions where the model is very sure the null hypothesis cannot be rejected (shrinkage by the ridge regression model). As a result, many foldchanges will be zero and the density plot for MSqRob may look like a spike instead of the expected Gaussian shape observed in other models*



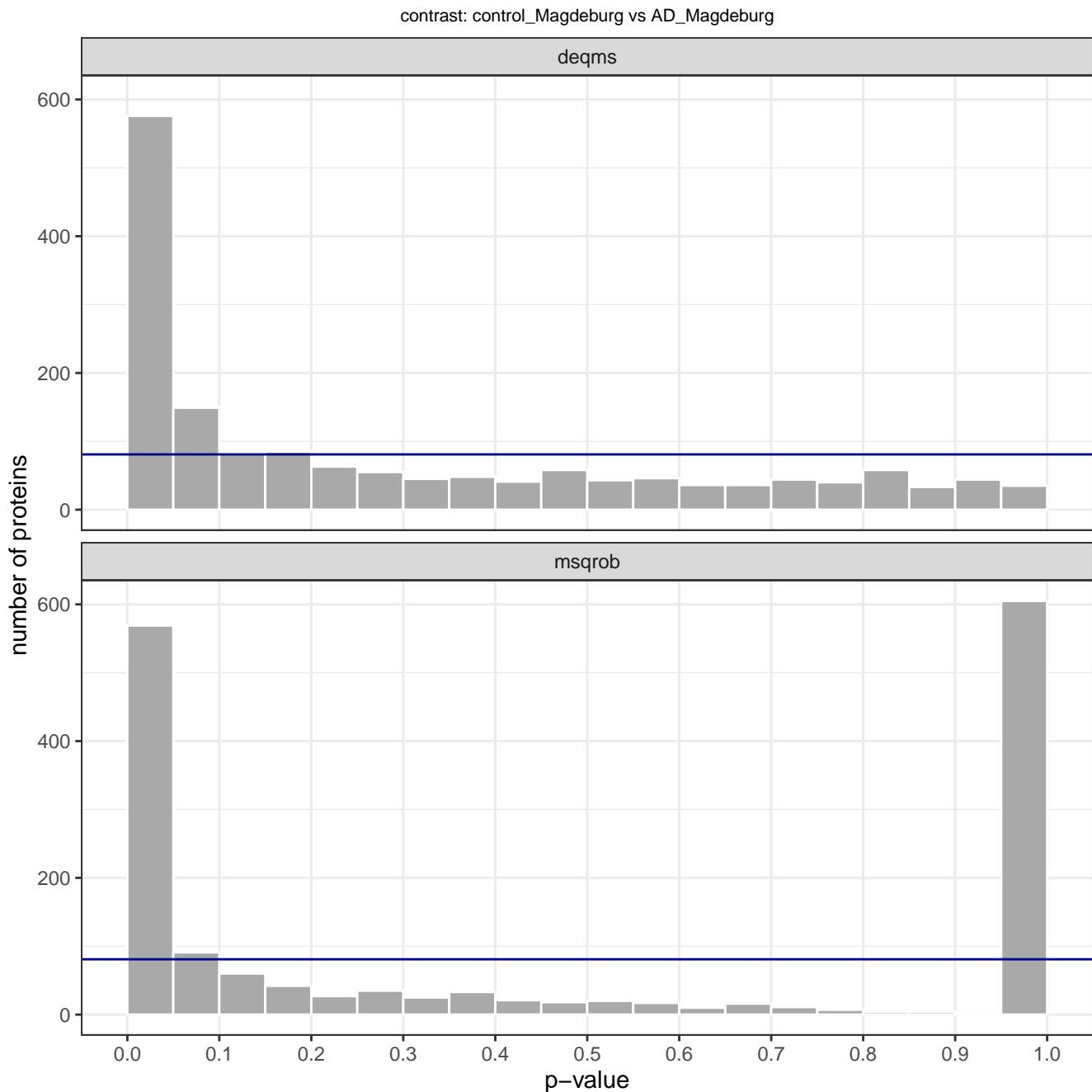
### 2.2.3 p-value distribution

Histogram of p-values computed by differential expression analysis algorithms, as-is, for quality-control inspection. The horizontal line indicates the expected counts assuming a uniform distribution (total number of p-values divided by number of histogram bins)

See further: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6164648/>

See further: <http://varianceexplained.org/statistics/interpreting-pvalue-histogram/>

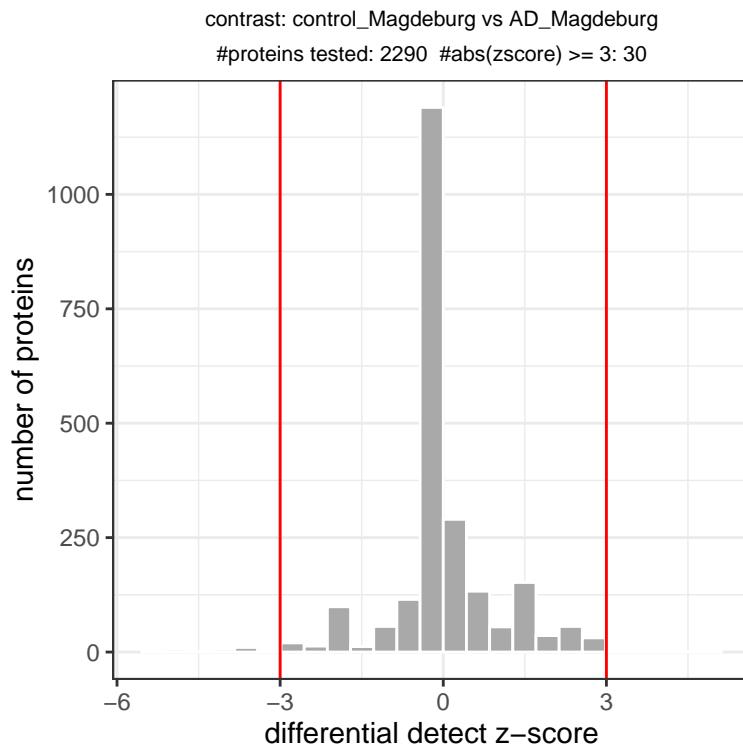
*note; the MSqRob and MS-Empire models often yield p-value distributions that show a large peak at p-value 1, these are typically proteins with estimated log foldchanges at/near zero where these models are very sure the null hypothesis cannot be rejected*



## 2.2.4 differential detect

Some proteins may not have peptides with sufficient data points over samples to be used for differential expression analysis (depending on the user-defined filtering criteria in how many replicates peptides should be observed), but do show a strong difference in the number of detected peptides between sample groups. In some proteomics experimental designs, for example a wildtype-knockout APMS study, those are interesting proteins. The DEA based on peptide abundance values (volcano plots above) are the main result for differential testing in MS-DAP but as a situationally useful tool MS-DAP also includes a ‘protein detection’ z-score, based on the number of times a peptide for each protein was detected per sample group (/experimental condition), as an alternative means of differential testing.

Below figure shows the distribution of these scores with thresholds at 3 std. Both the z-scores and the counts these are based upon are available in the statistical result Excel table.



## 2.3 control\_Sweden vs AD\_Sweden

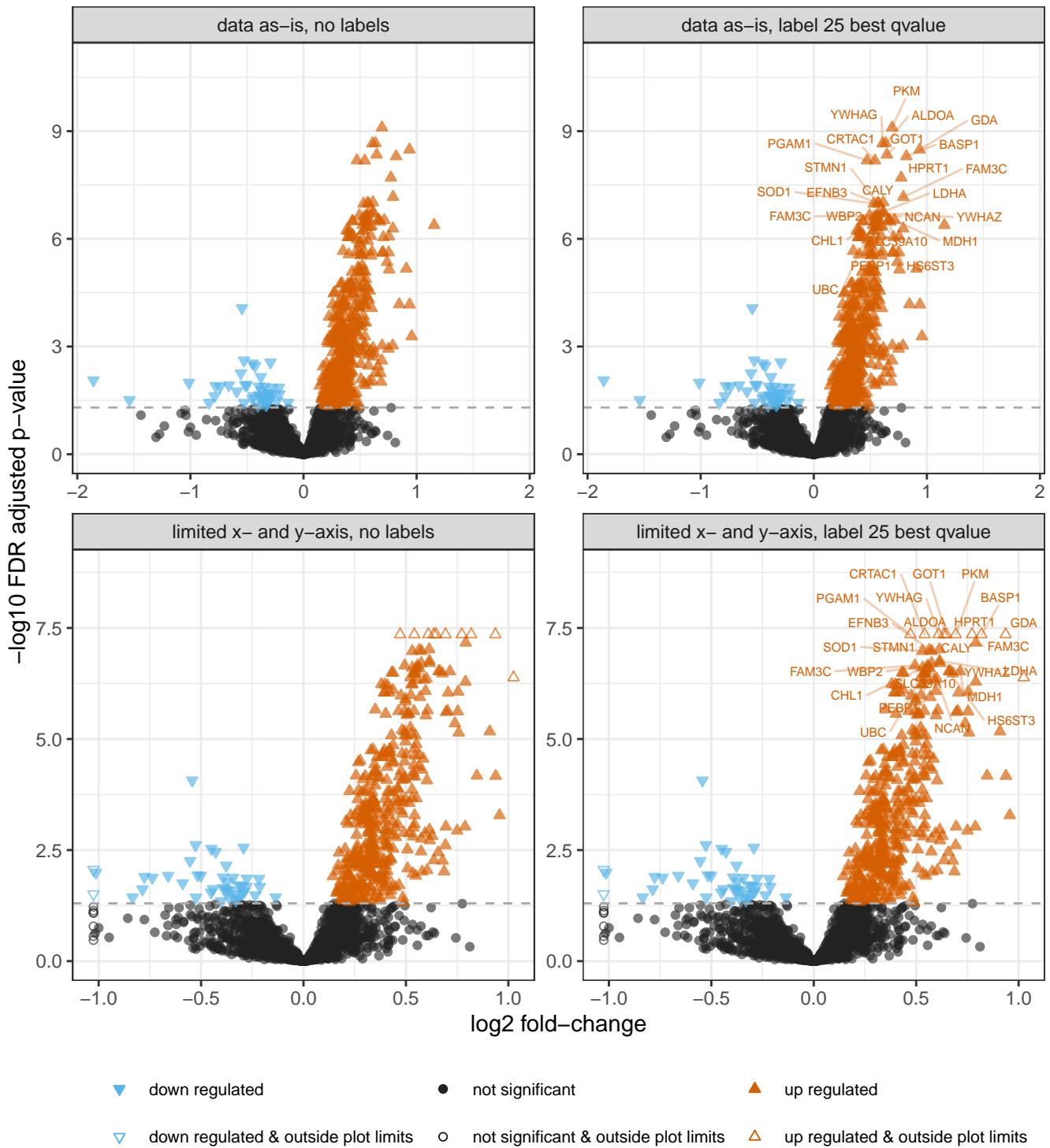
- **user setting:** using ‘filter by contrast’ peptide filtering approach
- 14272 peptides in 1732 proteins remain in the current contrast after peptide filters and are used for the statistical analysis in this section
- qvalue threshold: 0.05
- log2 foldchange threshold: 0

### 2.3.1 volcano

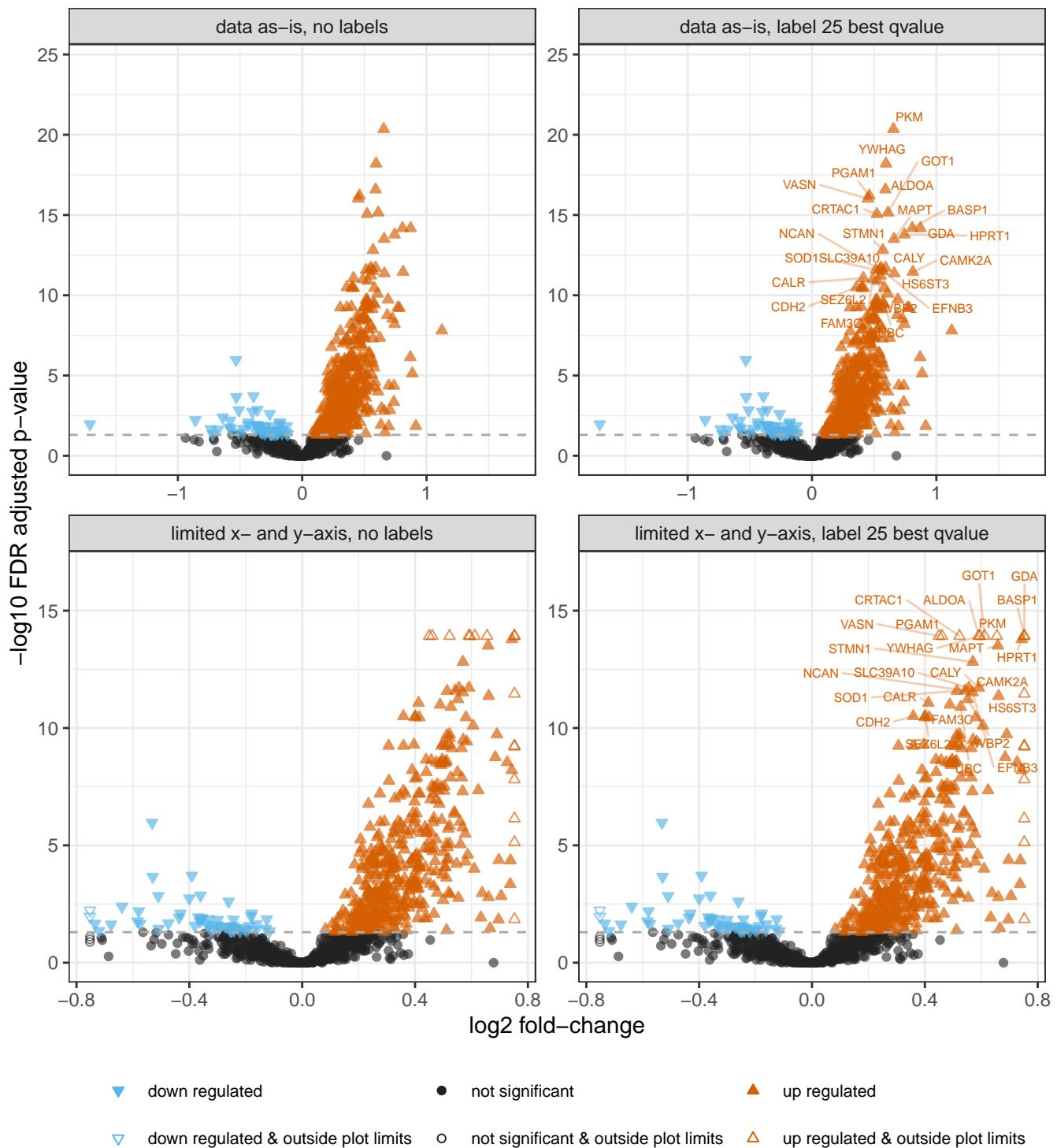
The plot title shows the statistical model and contrast (sample groups in the comparison). Left- and right-side figure panels on each row represent the same figure without and with labels for the 25 proteins with lowest p-value.

Bottom figure panels have limited x- and y-axis. For datasets with a small number of strong outliers in p-value or fold-change, which may have a profound effect on the plot scales, this allows inspection of the remainder of the volcano plot without disproportionate influence by ‘extreme’ values.

Labels for proteins that are more than 12 characters long are truncated for visual clarity (indicated by trailing ...). For protein identifiers that are ambiguous, e.g. a protein-group with assigned genes “gene1a;gene1b”, only the first label/ID is shown for visual clarity (indicated by trailing \*).



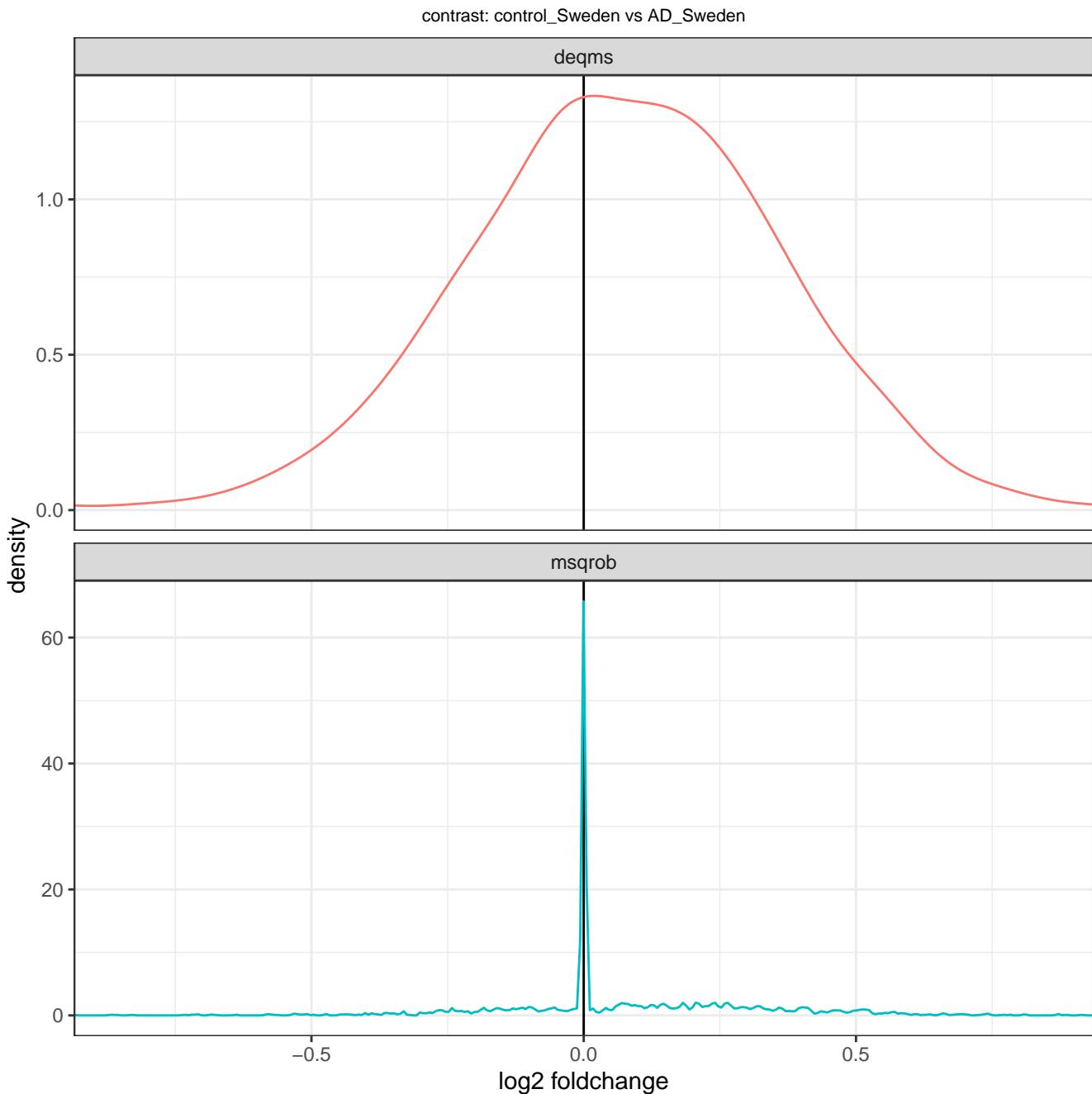
msqrob @ contrast: control\_Sweden vs AD\_Sweden



### 2.3.2 foldchange distribution

Distributions of estimated foldchanges produced by the statistical models. If the mode is far from 0, consider alternative normalization strategies. Do note the scale on the x-axis, for some experiments the foldchanges are very low which in turn may exaggerate this figure.

*note; the MSqRob model tends to assign zero (log)foldchange for proteins with minor difference between conditions where the model is very sure the null hypothesis cannot be rejected (shrinkage by the ridge regression model). As a result, many foldchanges will be zero and the density plot for MSqRob may look like a spike instead of the expected Gaussian shape observed in other models*



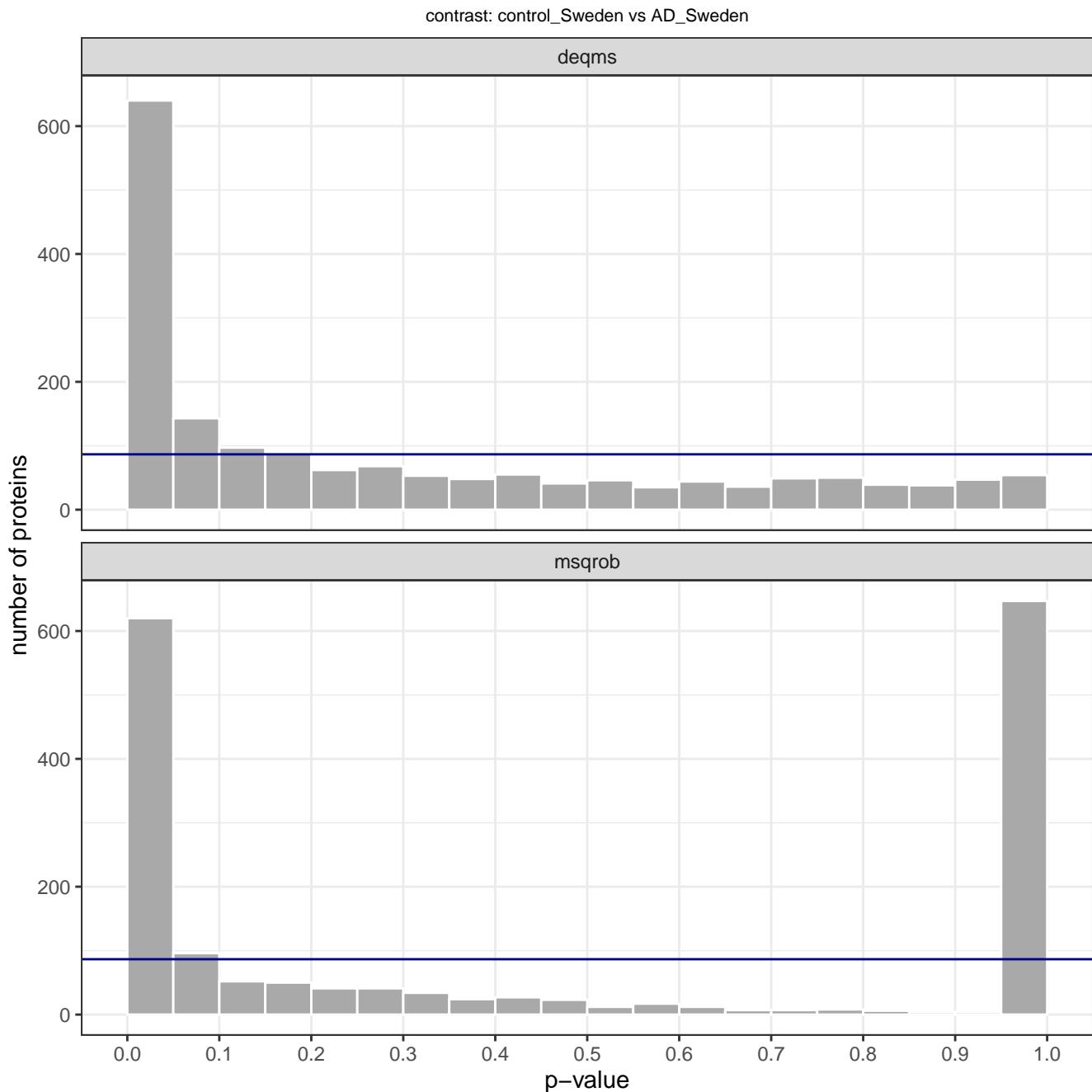
### 2.3.3 p-value distribution

Histogram of p-values computed by differential expression analysis algorithms, as-is, for quality-control inspection. The horizontal line indicates the expected counts assuming a uniform distribution (total number of p-values divided by number of histogram bins)

See further: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6164648/>

See further: <http://varianceexplained.org/statistics/interpreting-pvalue-histogram/>

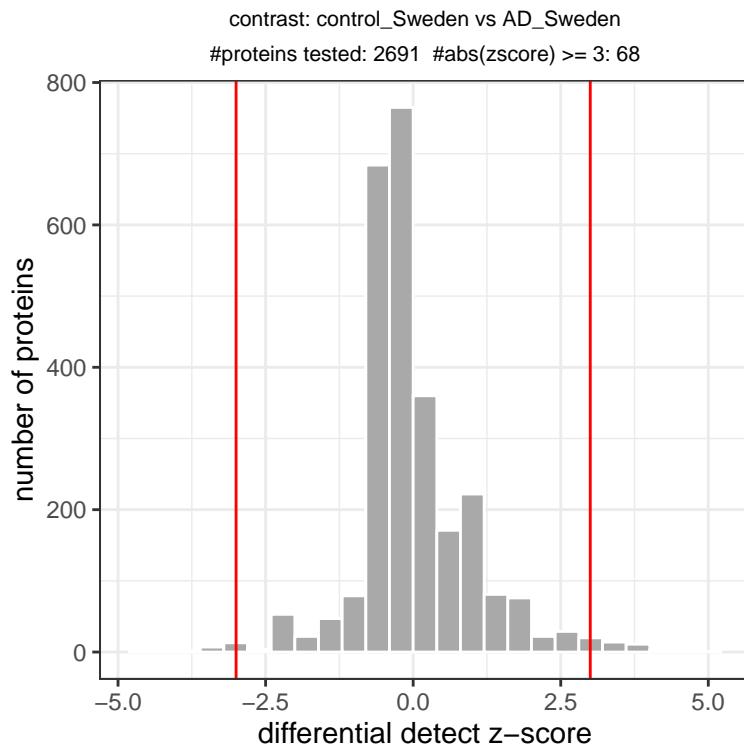
*note; the MSqRob and MS-Empire models often yield p-value distributions that show a large peak at p-value 1, these are typically proteins with estimated log foldchanges at/near zero where these models are very sure the null hypothesis cannot be rejected*



#### 2.3.4 differential detect

Some proteins may not have peptides with sufficient data points over samples to be used for differential expression analysis (depending on the user-defined filtering criteria in how many replicates peptides should be observed), but do show a strong difference in the number of detected peptides between sample groups. In some proteomics experimental designs, for example a wildtype-knockout APMS study, those are interesting proteins. The DEA based on peptide abundance values (volcano plots above) are the main result for differential testing in MS-DAP but as a situationally useful tool MS-DAP also includes a ‘protein detection’ z-score, based on the number of times a peptide for each protein was detected per sample group (/experimental condition), as an alternative means of differential testing.

Below figure shows the distribution of these scores with thresholds at 3 std. Both the z-scores and the counts these are based upon are available in the statistical result Excel table.



## 2.4 control\_Berlin,control\_Magdeburg... vs AD\_Berlin,AD\_Magdeburg,AD\_Sweden

- **user setting:** using ‘filter by contrast’ peptide filtering approach
- 11824 peptides in 1445 proteins remain in the current contrast after peptide filters and are used for the statistical analysis in this section
- qvalue threshold: 0.05
- log2 foldchange threshold: 0

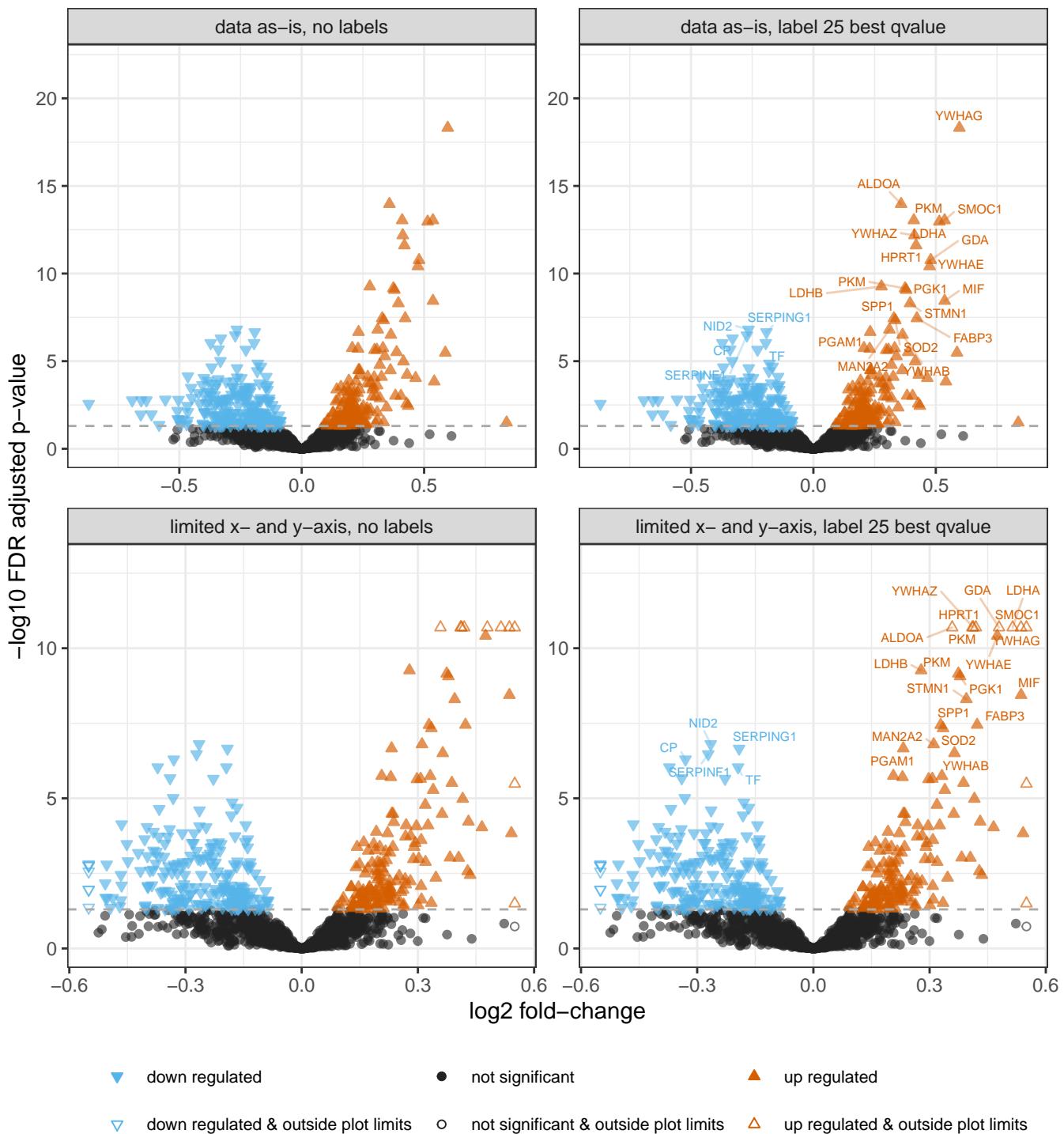
### 2.4.1 volcano

The plot title shows the statistical model and contrast (sample groups in the comparison). Left- and right-side figure panels on each row represent the same figure without and with labels for the 25 proteins with lowest p-value.

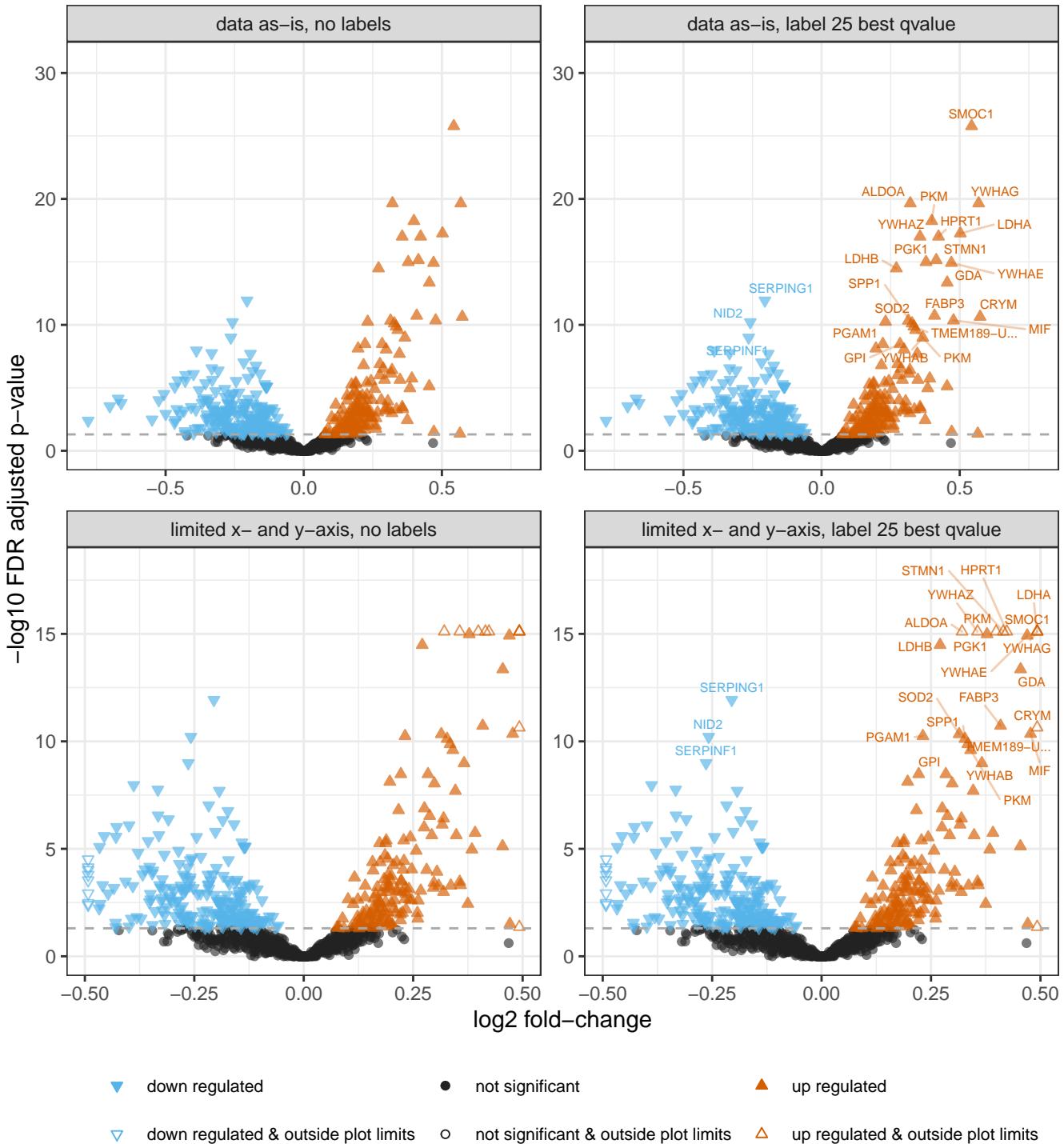
Bottom figure panels have limited x- and y-axis. For datasets with a small number of strong outliers in p-value or fold-change, which may have a profound effect on the plot scales, this allows inspection of the remainder of the volcano plot without disproportionate influence by ‘extreme’ values.

Labels for proteins that are more than 12 characters long are truncated for visual clarity (indicated by trailing ...). For protein identifiers that are ambiguous, e.g. a protein-group with assigned genes “gene1a;gene1b”, only the first label/ID is shown for visual clarity (indicated by trailing \*).

deqms @ contrast: control\_Berlin,control\_Magdeburg,control\_Sweden vs AD\_Berlin,AD\_Magdeburg,AD\_Sweden  
 user-specified random variables added to regression model: cohort



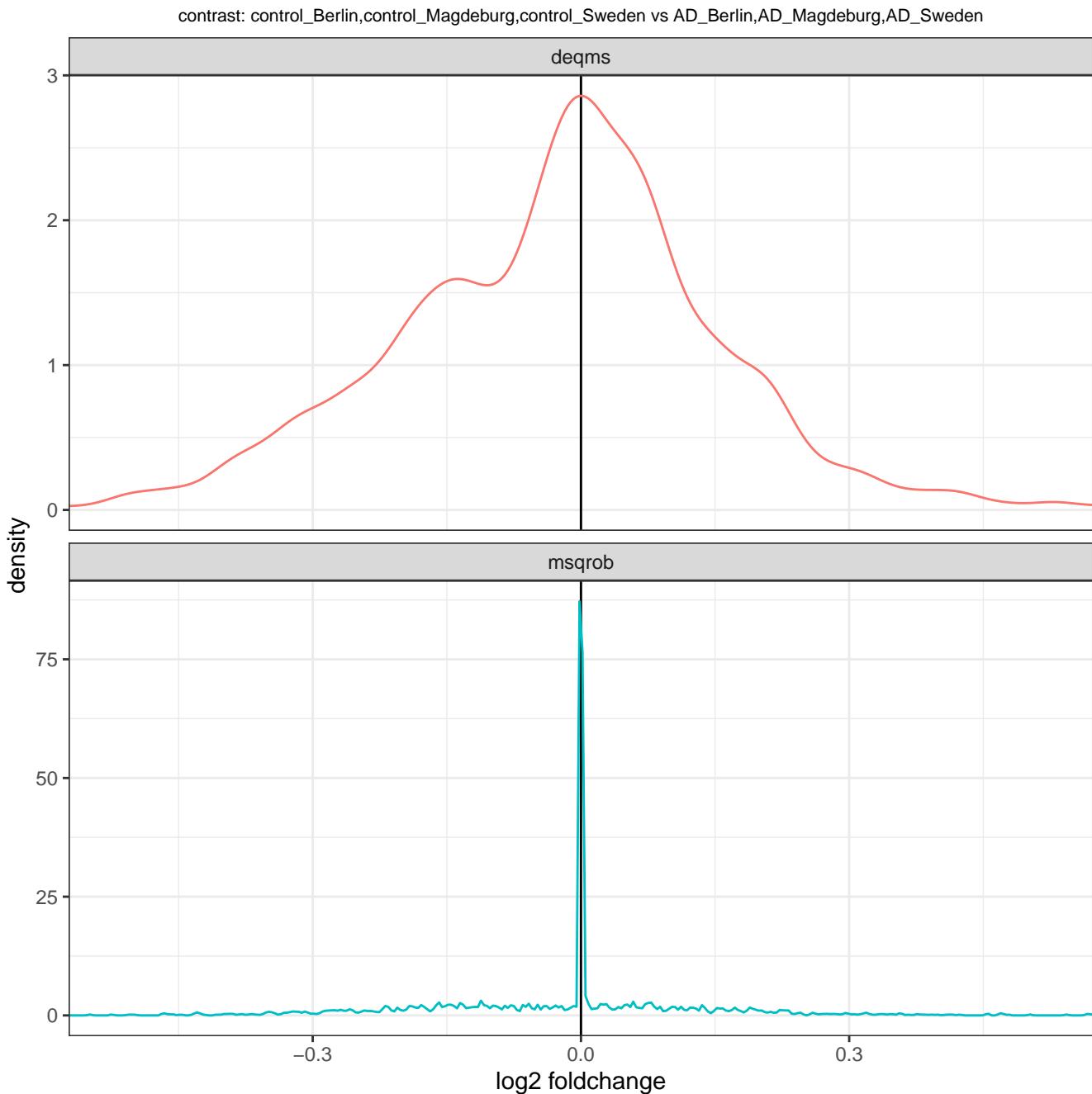
msqrob @ contrast: control\_Berlin,control\_Magdeburg,control\_Sweden vs AD\_Berlin,AD\_Magdeburg,AD\_Sweden  
user-specified random variables added to regression model: cohort



## 2.4.2 foldchange distribution

Distributions of estimated foldchanges produced by the statistical models. If the mode is far from 0, consider alternative normalization strategies. Do note the scale on the x-axis, for some experiments the foldchanges are very low which in turn may exaggerate this figure.

*note; the MSqRob model tends to assign zero (log)foldchange for proteins with minor difference between conditions where the model is very sure the null hypothesis cannot be rejected (shrinkage by the ridge regression model). As a result, many foldchanges will be zero and the density plot for MSqRob may look like a spike instead of the expected Gaussian shape observed in other models*



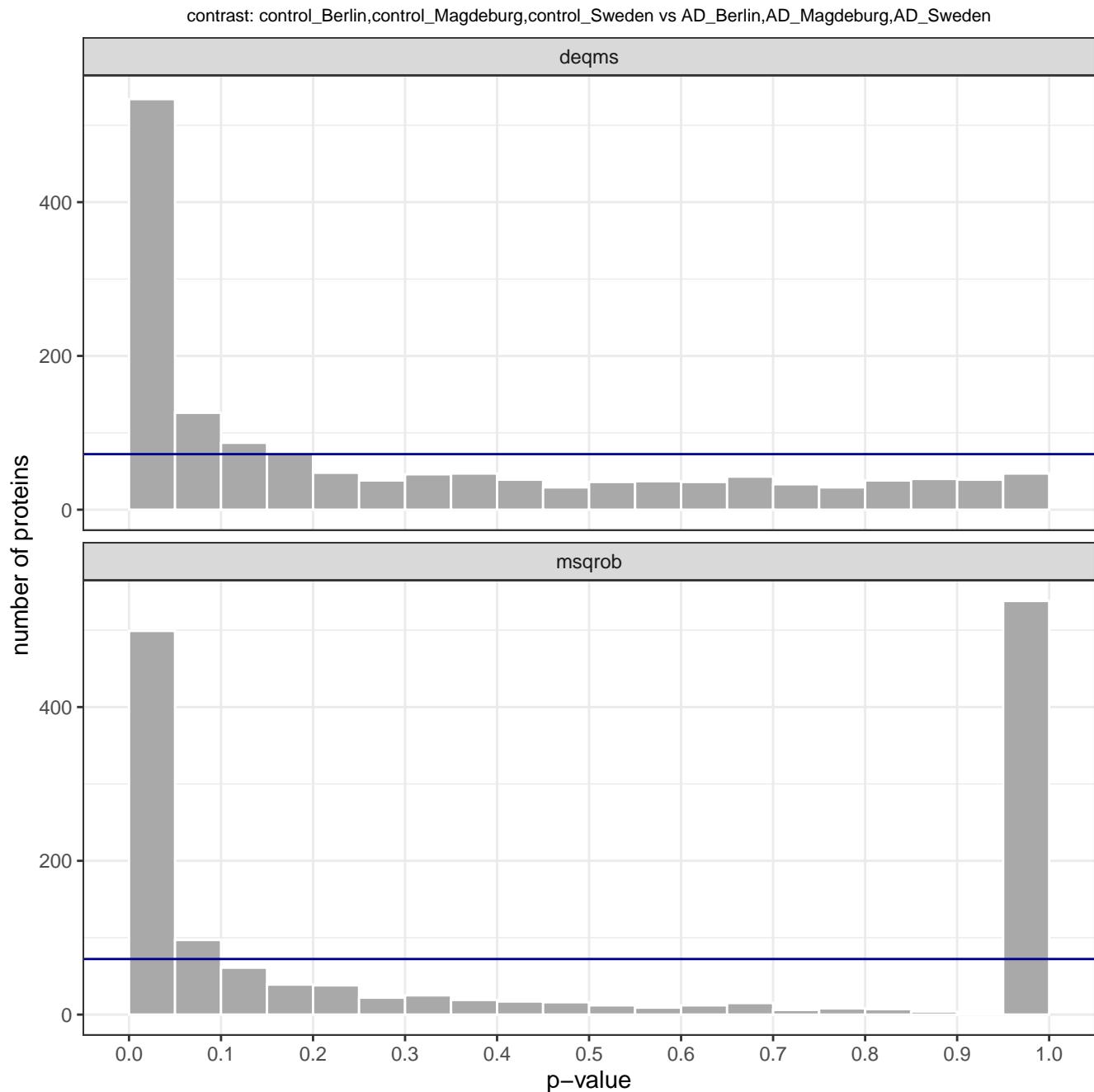
### 2.4.3 p-value distribution

Histogram of p-values computed by differential expression analysis algorithms, as-is, for quality-control inspection. The horizontal line indicates the expected counts assuming a uniform distribution (total number of p-values divided by number of histogram bins)

See further: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6164648/>

See further: <http://varianceexplained.org/statistics/interpreting-pvalue-histogram/>

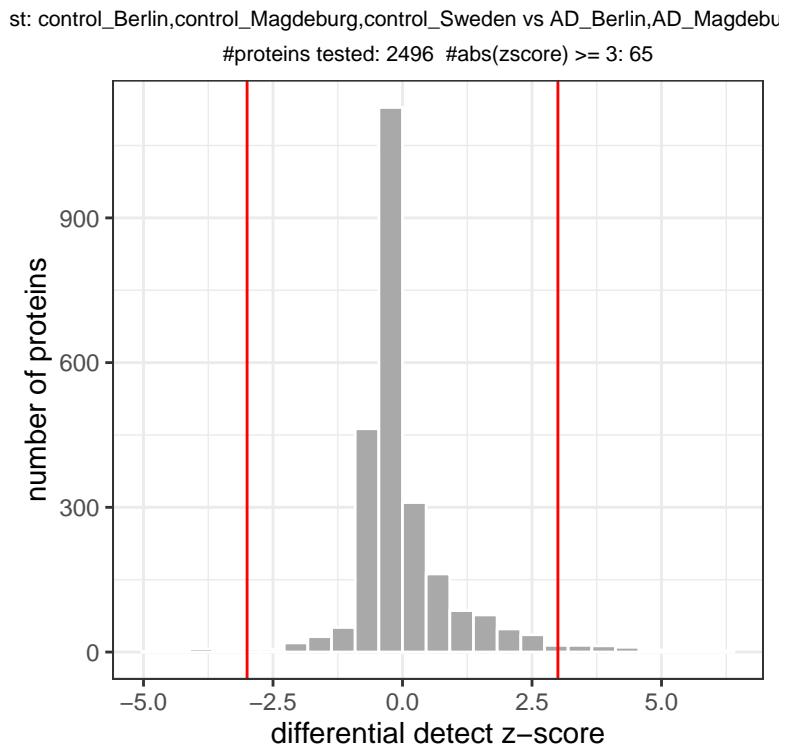
*note; the MSqRob and MS-Empire models often yield p-value distributions that show a large peak at p-value 1, these are typically proteins with estimated log foldchanges at/near zero where these models are very sure the null hypothesis cannot be rejected*



#### 2.4.4 differential detect

Some proteins may not have peptides with sufficient data points over samples to be used for differential expression analysis (depending on the user-defined filtering criteria in how many replicates peptides should be observed), but do show a strong difference in the number of detected peptides between sample groups. In some proteomics experimental designs, for example a wildtype-knockout APMS study, those are interesting proteins. The DEA based on peptide abundance values (volcano plots above) are the main result for differential testing in MS-DAP but as a situationally useful tool MS-DAP also includes a ‘protein detection’ z-score, based on the number of times a peptide for each protein was detected per sample group (/experimental condition), as an alternative means of differential testing.

Below figure shows the distribution of these scores with thresholds at 3 std. Both the z-scores and the counts these are based upon are available in the statistical result Excel table.



### 3 Summary of differential testing

Differential Expression Analysis: number of proteins found statistically significant.

contrast		algorithm	#test	#hits	top10 significant
control_Berlin vs AD_Berlin		deqms	1662	17	ywhag, smoc1, ywhaz, ywhae, aldoa, ldha, pgk1, gda, hpert1, enol
control_Berlin vs AD_Berlin		msqrob	1662	31	ywhag, smoc1, ywhae, ywhaz, aldoa, pgk1, ldha, hpert1, gda, ptpa
control_Magdeburg	vs	deqms	1617	384	mif, ldha, dld, serpinf1, ltbp1, camk2b, rims2, dynll1;..., clic4, pgam1
control_Magdeburg	vs	msqrob	1617	411	mif, dynll1;..., camk2b, gpi, ldha, serpinf1, pgam1, clic4, nid2, ltbp1
AD_Magdeburg					
control_Sweden	vs	deqms	1732	453	pkm, aldoa, ywhag, gda, got1, basp1, pgam1, crtac1, hpert1, fam3c
AD_Sweden					
control_Sweden	vs	msqrob	1732	492	pkm, ywhag, aldoa, pgam1, vasn, got1, crtac1, basp1, gda, hpert1
AD_Sweden					
control_Berlin,...	vs	deqms	1445	380	ywhag, aldoa, smoc1, pkm, ldha, ywhaz, hpert1, gda, ywhae, ldhb
AD_Berlin,AD_Ma...					
control_Berlin,...	vs	msqrob	1445	379	smoc1, ywhag, aldoa, pkm, ldha, hpert1, ywhaz, stmn1, pgk1, ywhae
AD_Berlin,AD_Ma...					

Differential Detection: prioritize proteins with more peptide detections in some group. A simple metric to complement results from DEA, which is the main result (eg; consider proteins with too few data points for DEA).

contrast		#proteins tested	#abs(zscore) >= 3	top10
control_Berlin vs AD_Berlin		2369	74	park7, grm4, acly, arl3, sdk2, ptprc, sema5a, mgat4a, cct2, psmd1
control_Magdeburg	vs	2290	30	crhbp, ppp3ca, wars, btf3, stx1b, mydgf, hspa4, brinp1, pcyox1, slitrk6
AD_Magdeburg				
control_Sweden	vs	2691	68	hnrrnpk, ddx6, rps10-n..., tomm22, rab14, ssrp1, glul, dlat, hist1h2..., ccsms1
AD_Sweden				
control_Berlin,...	vs	2496	65	tubb, park7, plppr4, smarce1, cdon, glce, dlat, gsta4, rbmxl1;..., sqstm1
AD_Berlin,AD_Ma...				

## 4 log

```
[info] reading Spectronaut report...
[info] 22557/31469 precursors remain after selecting the 'best' precursor for each modified sequence
[info] 6882/6882 protein accessions and 2771/2771 protein groups were mapped to provided fasta file(s)
[info] contrast: control_Berlin vs AD_Berlin
[info] contrast: control_Magdeburg vs AD_Magdeburg
[info] contrast: control_Sweden vs AD_Sweden
[info] contrast: control_Berlin,control_Magdeburg,control_Sweden vs AD_Berlin,AD_Magdeburg,AD_Sweden
[info] using 23 threads for multiprocessing
[progress] caching filter data took 2.5 minutes
[progress] peptide to protein rollup with MaxLFQ (implementation: iq) took 1 seconds
[progress] peptide to protein rollup with MaxLFQ (implementation: iq) took 2 seconds
[progress] peptide to protein rollup with MaxLFQ (implementation: iq) took 1 seconds
[progress] peptide to protein rollup with MaxLFQ (implementation: iq) took 1 seconds
[progress] peptide to protein rollup with MaxLFQ (implementation: iq) took 1 seconds
[progress] peptide to protein rollup with MaxLFQ (implementation: iq) took 2 seconds
[progress] peptide filtering and normalization took 4.1 minutes
[info] peptide to protein rollup strategy: maxlfq
[info] differential expression analysis for contrast: control_Berlin vs AD_Berlin
[info] using data from peptide filter: filter by contrast
[info] random variables that are not applicable for current contrast due to lack of unique values (compared to sample groups/condition): cohort
[progress] peptide to protein rollup with MaxLFQ (implementation: iq) took 1 seconds
[progress] DEqMS took 1 seconds
[info] msqrob linear regression formulas (these are prioritized. eg; if a model fit fails due to lack of data, the next formula is used); expression ~ (1 | condition) + (1 | sample_id) + (1 | peptide_id) , expression ~ (1 | condition)
[progress] msqrob took 6.5 minutes
[info] differential expression analysis for contrast: control_Magdeburg vs AD_Magdeburg
[info] using data from peptide filter: filter by contrast
[info] random variables that are not applicable for current contrast due to lack of unique values (compared to sample groups/condition): cohort
[progress] peptide to protein rollup with MaxLFQ (implementation: iq) took 1 seconds
[progress] DEqMS took 1 seconds
[info] msqrob linear regression formulas (these are prioritized. eg; if a model fit fails due to lack of data, the next formula is used); expression ~ (1 | condition) + (1 | sample_id) + (1 | peptide_id) , expression ~ (1 | condition)
[progress] msqrob took 5.7 minutes
[info] differential expression analysis for contrast: control_Sweden vs AD_Sweden
[info] using data from peptide filter: filter by contrast
[info] random variables that are not applicable for current contrast due to lack of unique values (compared to sample groups/condition): cohort
[progress] peptide to protein rollup with MaxLFQ (implementation: iq) took 1 seconds
[progress] DEqMS took 1 seconds
[info] msqrob linear regression formulas (these are prioritized. eg; if a model fit fails due to lack of data, the next formula is used); expression ~ (1 | condition) + (1 | sample_id) + (1 | peptide_id) , expression ~ (1 | condition)
[progress] msqrob took 5.6 minutes
[info] differential expression analysis for contrast: control_Berlin,control_Magdeburg,control_Sweden vs AD_Berlin,AD_Magdeburg,AD_Sweden
[info] using data from peptide filter: filter by contrast
[info] random variables used in current contrast: cohort
[progress] peptide to protein rollup with MaxLFQ (implementation: iq) took 1 seconds
```

```
[progress] DEqMS took 1 seconds
[info] msqrob linear regression formulas (these are prioritized. eg; if a model fit fails due to lack of data, the next formula is used); expression ~ (1 | condition) + (1 | sample_id) + (1 | peptide_id) + (1 | cohort) , expression ~ (1 | condition) + (1 | cohort) , expression ~ (1 | condition)
[progress] msqrob took 10.3 minutes
[info] differential detection analysis: min_samples_observed=3
[progress] peptide to protein rollup with MaxLFQ (implementation: iq) took 1 seconds
[progress] peptide to protein rollup with MaxLFQ (implementation: iq) took 1 seconds
[progress] peptide to protein rollup with MaxLFQ (implementation: iq) took 1 seconds
[progress] peptide to protein rollup with MaxLFQ (implementation: iq) took 1 seconds
[progress] peptide to protein rollup with MaxLFQ (implementation: iq) took 1 seconds
[progress] peptide to protein rollup with MaxLFQ (implementation: iq) took 1 seconds
[progress] peptide to protein rollup with MaxLFQ (implementation: iq) took 1 seconds
[progress] peptide to protein rollup with MaxLFQ (implementation: iq) took 1 seconds
[progress] peptide to protein rollup with MaxLFQ (implementation: iq) took 1 seconds
[progress] creating PDF report...
[progress] report: constructing plots specific for each contrast
[progress] report: rendering report (this may take a while depending on dataset size)
[progress] RT plots: preparing data took 20 seconds
[progress] RT plots: creating plots took 7 seconds
[info] skipping within-group foldchange plots for sample group 'unknown_Kiel', require at least 2 replicates
[info] skipping within-group foldchange plots for sample group 'unknown_Sweden', require at least 2 replicates
[info] No data available for CoV leave-one-out computation in sample group 'unknown_Kiel', skipping plots
[info] No data available for CoV leave-one-out computation in sample group 'unknown_Sweden', skipping plots
[progress] leave-one-out CoV plot computations took 8 seconds
[info] no CoV computation for sample group 'unknown_Kiel', require at least 3 replicates
[info] no CoV computation for sample group 'unknown_Magdeburg', require at least 3 replicates
[info] no CoV computation for sample group 'unknown_Sweden', require at least 3 replicates
[progress] peptide to protein rollup with MaxLFQ (implementation: iq) took 1 seconds
[info] no CoV computation for sample group 'unknown_Kiel', require at least 3 replicates
[info] no CoV computation for sample group 'unknown_Magdeburg', require at least 3 replicates
[info] no CoV computation for sample group 'unknown_Sweden', require at least 3 replicates
```

## 5 R command history

This shows the history commands from your R script that starts this pipeline, thereby automatically documenting the parameters/settings used. All lines of executed code since (last) importing data using this R package are shown.

### Using this feature

Do not use RStudio's `source` option to execute our pipeline since it will only write `source(...yourscript.R)` to the session history, and consequentially that is all you see in this 'code log'. Instead, select all lines in your script (`control + A`) and then "run" the selected code (either click the run button in RStudio, or use `control + enter`). All lines shown in this section are the same as shown in the RStudio 'History' pane (a tab on the top-right of its UI).

```
dataset = import_dataset_spectronaut(
  filename = "C:/DATA/PMID32485097/20200914_185722_SNA-fSN1.1_CLib1.1 with deamidation_SNA-SER1.1 BGS default_all final raw files Zet Ma
  confidence_threshold = 0.05,
  use_normalized_intensities = FALSE,
  use_irt = TRUE, do_plot = TRUE
)
dataset = import_fasta(dataset,
  files = c(
    "C:/fasta/UniProt_2018-05/UP000005640_9606.fasta",
    "C:/fasta/UniProt_2018-05/UP000005640_9606_additional.fasta",
    "C:/fasta/UniProt_2017-06/UP000005640_9606.fasta",
    "C:/fasta/UniProt_2017-06/UP000005640_9606_additional.fasta"
  )
)
dataset = import_sample_metadata(dataset,
  filename = "C:/DATA/PMID32485097/sample_metadata.xlsx"
)
dataset = setup_contrasts(dataset,
  contrast_list = list(
    c(
      "control_Berlin",
      "AD_Berlin"
    ),
    c(
      "control_Magdeburg",
      "AD_Magdeburg"
    ),
    c(
      "control_Sweden",
      "AD_Sweden"
    ),
    list(
      c(
        "control_Berlin",
        "control_Magdeburg",
        "control_Sweden"
      ),
      c(
        "AD_Berlin",
        "AD_Magdeburg",
        "AD_Sweden"
      )
    )
  ),
  random_variables = "cohort"
)
dataset = analysis_quickstart(dataset,
  filter_min_detect = 8,
  filter_by_contrast = TRUE,
  filter_topn_peptides = 0,
  filter_min_peptide_per_prot = 1,
  norm_algorithm = c(
    "vsn",
    "modebetween_protein"
  ),
  dea_algorithm = c(
    "deqms",
    "deqms"
  )
```

```
"msqrob"  
) , dea_qvalue_threshold = 0.05,  
output_qc_report = TRUE,  
output_dir = "C:/DATA/PMID32485097/",  
dump_all_data = TRUE,  
output_within_timestamped_subdirectory = TRUE  
)
```

## 6 R session info

The computer system and versioning of all R packages used to run this analysis are shown below to facilitate, in combination with the previous section, reproducibility.

setting	value
version	R version 3.6.3 (2020-02-29)
os	Windows 10 x64
system	x86_64, mingw32
ui	RStudio
language	(EN)
collate	English_United States.1252
ctype	English_United States.1252
tz	Europe/Berlin
date	2021-05-23

*System*

package	loadedversion	source
dplyr	1.0.0	CRAN (R 3.6.3)
ggplot2	3.3.2	CRAN (R 3.6.3)
msdap	0.2.8.1	local
rlang	0.4.6	CRAN (R 3.6.3)
testthat	2.3.2	CRAN (R 3.6.3)
tibble	3.0.1	CRAN (R 3.6.3)
tidyverse	1.1.0	CRAN (R 3.6.3)

*Attached packages*

package	loadedversion	source
abind	1.4-5	CRAN (R 3.6.0)
affy	1.64.0	Bioconductor
affyio	1.56.0	Bioconductor
AnnotationDbi	1.48.0	Bioconductor
arrangements	1.1.8	CRAN (R 3.6.3)
askpass	1.1	CRAN (R 3.6.3)
assertthat	0.2.1	CRAN (R 3.6.3)
backports	1.1.7	CRAN (R 3.6.3)
bigassertr	0.1.4	CRAN (R 3.6.3)
bigparallelr	0.3.1	CRAN (R 3.6.3)
bigstatsr	1.5.1	CRAN (R 3.6.3)
Biobase	2.46.0	Bioconductor
BiocFileCache	1.10.2	Bioconductor
BiocGenerics	0.32.0	Bioconductor
BiocManager	1.30.10	CRAN (R 3.6.3)
BiocParallel	1.20.1	Bioconductor
biomaRt	2.45.8	Github (grimbough/biomaRt@61c3ee0)
bit	1.1-15.2	CRAN (R 3.6.2)
bit64	0.9-7	CRAN (R 3.6.2)
bitops	1.0-6	CRAN (R 3.6.0)
blob	1.2.1	CRAN (R 3.6.3)
boot	1.3-24	CRAN (R 3.6.3)
broom	0.5.6	CRAN (R 3.6.3)
callr	3.4.3	CRAN (R 3.6.3)
car	3.0-8	CRAN (R 3.6.3)
carData	3.0-4	CRAN (R 3.6.3)
caTools	1.18.0	CRAN (R 3.6.3)
cellranger	1.1.0	CRAN (R 3.6.3)
cli	2.0.2	CRAN (R 3.6.3)
codetools	0.2-16	CRAN (R 3.6.3)
colorRamps	2.3	CRAN (R 3.6.0)
colorspace	1.4-1	CRAN (R 3.6.3)
cowplot	1.0.0	CRAN (R 3.6.3)
crayon	1.3.4	CRAN (R 3.6.3)
curl	4.3	CRAN (R 3.6.3)
data.table	1.12.8	CRAN (R 3.6.3)
DBI	1.1.0	CRAN (R 3.6.3)
dbplyr	1.4.4	CRAN (R 3.6.3)
DEqMS	1.4.0	Bioconductor
desc	1.2.0	CRAN (R 3.6.3)
devtools	2.3.0	CRAN (R 3.6.3)
diann	1.0.1	Github (vdemichev/diann-rpackage@af538f6)
digest	0.6.25	CRAN (R 3.6.3)
doParallel	1.0.15	CRAN (R 3.6.3)
ellipsis	0.3.1	CRAN (R 3.6.3)
eulerr	6.1.0	CRAN (R 3.6.3)
evaluate	0.14	CRAN (R 3.6.3)
fansi	0.4.1	CRAN (R 3.6.3)
farver	2.0.3	CRAN (R 3.6.3)
flock	0.7	CRAN (R 3.6.3)

package	loadedversion	source
forcats	0.5.0	CRAN (R 3.6.3)
foreach	1.5.0	CRAN (R 3.6.3)
foreign	0.8-75	CRAN (R 3.6.3)
formatR	1.7	CRAN (R 3.6.3)
fs	1.4.1	CRAN (R 3.6.3)
gdata	2.18.0	CRAN (R 3.6.3)
generics	0.1.0	CRAN (R 3.6.3)
ggpubr	0.4.0	CRAN (R 3.6.3)
ggrepel	0.8.2	CRAN (R 3.6.3)
ggsignif	0.6.0	CRAN (R 3.6.3)
glue	1.4.1	CRAN (R 3.6.3)
gmp	0.6-0	CRAN (R 3.6.3)
gplots	3.0.4	CRAN (R 3.6.3)
gridExtra	2.3	CRAN (R 3.6.3)
gtable	0.3.0	CRAN (R 3.6.3)
gtools	3.8.2	CRAN (R 3.6.3)
haven	2.3.1	CRAN (R 3.6.3)
hms	0.5.3	CRAN (R 3.6.3)
htmltools	0.5.0	CRAN (R 3.6.3)
httr	1.4.1	CRAN (R 3.6.3)
impute	1.60.0	Bioconductor
iq	1.9	CRAN (R 3.6.3)
IRanges	2.20.2	Bioconductor
iterators	1.0.12	CRAN (R 3.6.3)
KernSmooth	2.23-16	CRAN (R 3.6.3)
knitr	1.29	CRAN (R 3.6.3)
labeling	0.3	CRAN (R 3.6.0)
lattice	0.20-38	CRAN (R 3.6.3)
lifecycle	0.2.0	CRAN (R 3.6.3)
limma	3.42.2	Bioconductor
lme4	1.1-23	CRAN (R 3.6.3)
magrittr	1.5	CRAN (R 3.6.3)
MALDIquant	1.19.3	CRAN (R 3.6.3)
MASS	7.3-51.5	CRAN (R 3.6.3)
Matrix	1.2-18	CRAN (R 3.6.3)
matrixStats	0.56.0	CRAN (R 3.6.3)
memoise	1.1.0	CRAN (R 3.6.3)
mgcv	1.8-31	CRAN (R 3.6.3)
minqa	1.2.4	CRAN (R 3.6.3)
msEmpiRe	0.1.0	Github (zimmerlab/MS-EmpiRe@8a85757)
MSnbase	2.12.0	Bioconductor
munsell	0.5.0	CRAN (R 3.6.3)
mzID	1.24.0	Bioconductor
mzR	2.20.0	Bioconductor
ncdf4	1.17	CRAN (R 3.6.1)
nlme	3.1-144	CRAN (R 3.6.3)
nloptr	1.2.2.1	CRAN (R 3.6.3)
openssl	1.4.2	CRAN (R 3.6.3)
openxlsx	4.1.5	CRAN (R 3.6.3)
packrat	0.5.0	CRAN (R 3.6.3)

package	loadedversion	source
patchwork	1.0.1	CRAN (R 3.6.3)
pkrtest	0.4-8.6	CRAN (R 3.6.3)
pcaMethods	1.78.0	Bioconductor
pdftools	2.3.1	CRAN (R 3.6.3)
pillar	1.4.4	CRAN (R 3.6.3)
pkgbuild	1.0.8	CRAN (R 3.6.3)
pkgconfig	2.0.3	CRAN (R 3.6.3)
pkgload	1.1.0	CRAN (R 3.6.3)
plyr	1.8.6	CRAN (R 3.6.3)
preprocessCore	1.48.0	Bioconductor
prettyunits	1.1.1	CRAN (R 3.6.3)
pROC	1.16.2	CRAN (R 3.6.3)
processx	3.4.2	CRAN (R 3.6.3)
progress	1.2.2	CRAN (R 3.6.3)
ProtGenerics	1.18.0	Bioconductor
ps	1.3.3	CRAN (R 3.6.3)
purrr	0.3.4	CRAN (R 3.6.3)
qpdf	1.1	CRAN (R 3.6.3)
R.methodsS3	1.8.0	CRAN (R 3.6.2)
R.oo	1.23.0	CRAN (R 3.6.2)
R.utils	2.9.2	CRAN (R 3.6.3)
R6	2.4.1	CRAN (R 3.6.3)
rappdirs	0.3.1	CRAN (R 3.6.3)
RColorBrewer	1.1-2	CRAN (R 3.6.0)
Rcpp	1.0.4.6	CRAN (R 3.6.3)
RcppEigen	0.3.3.7.0	CRAN (R 3.6.3)
readr	1.3.1	CRAN (R 3.6.3)
readxl	1.3.1	CRAN (R 3.6.3)
remotes	2.1.1	CRAN (R 3.6.3)
reshape2	1.4.4	CRAN (R 3.6.3)
rio	0.5.16	CRAN (R 3.6.3)
rmarkdown	2.3	CRAN (R 3.6.3)
rprojroot	1.3-2	CRAN (R 3.6.3)
rsconnect	0.8.16	CRAN (R 3.6.3)
RSQLite	2.2.0	CRAN (R 3.6.3)
rstatix	0.6.0	CRAN (R 3.6.3)
rstudioapi	0.11	CRAN (R 3.6.3)
S4Vectors	0.24.4	Bioconductor
scales	1.1.1	CRAN (R 3.6.3)
sessioninfo	1.1.1	CRAN (R 3.6.3)
statmod	1.4.34	CRAN (R 3.6.3)
stringi	1.4.6	CRAN (R 3.6.2)
stringr	1.4.0	CRAN (R 3.6.3)
styler	1.3.2	CRAN (R 3.6.3)
tidyselect	1.1.0	CRAN (R 3.6.3)
tinytex	0.24	CRAN (R 3.6.3)
useThis	1.6.1	CRAN (R 3.6.3)
variancePartition	1.16.1	Bioconductor
vctrs	0.3.1	CRAN (R 3.6.3)
viridis	0.5.1	CRAN (R 3.6.3)

package	loadedversion	source
viridisLite	0.3.0	CRAN (R 3.6.3)
vsn	3.54.0	Bioconductor
withr	2.2.0	CRAN (R 3.6.3)
xfun	0.15	CRAN (R 3.6.3)
XML	3.99-0.3	CRAN (R 3.6.3)
xml2	1.3.2	CRAN (R 3.6.3)
xtable	1.8-4	CRAN (R 3.6.3)
yaml	2.2.1	CRAN (R 3.6.3)
zip	2.0.4	CRAN (R 3.6.3)
zlibbioc	1.32.0	Bioconductor

*Packages that are not attached*