Proteome-wide label-free quantification with MaxQuant

Jürgen Cox
Max Planck Institute of Biochemistry
July 2011

MaxQuant

Data acquisition

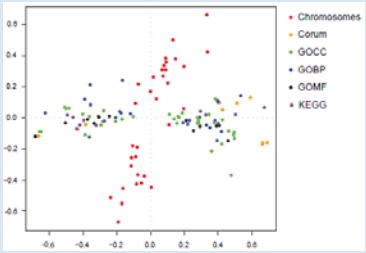
Raw data

MaxQuant

- Feature detection
- Initial Andromeda search
- Recalibration
- Main Andromeda search
- Consolidation / protein quantification

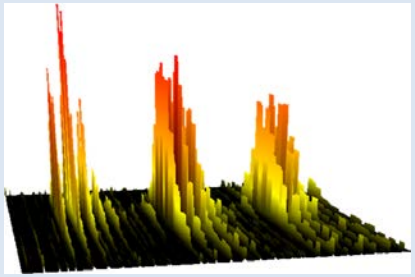
MQ output tables

Statistics & systems biology



Perseus

Inspection of raw data



Viewer

Supported input data

Labeling methods

- SILAC
- Label free
- Di-methyl
- 18O
- ICAT
- ICPL

Work in progress:

- iTRAQ

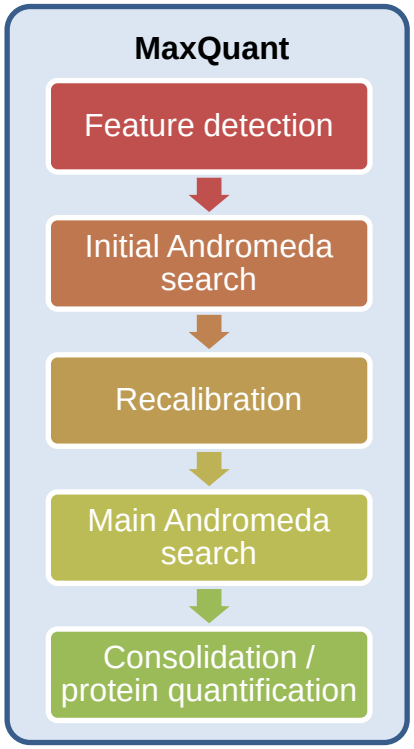
Mass spectrometers

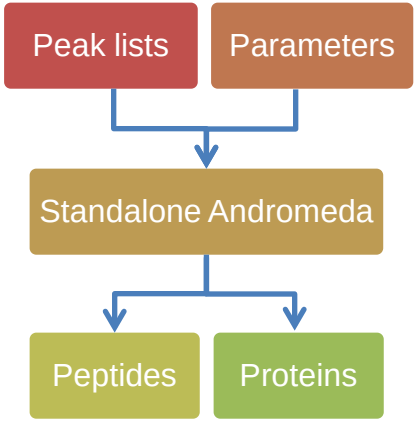
- Thermo Fisher Orbitrap and FT

Work in progress:

- SCIEX Triple TOF

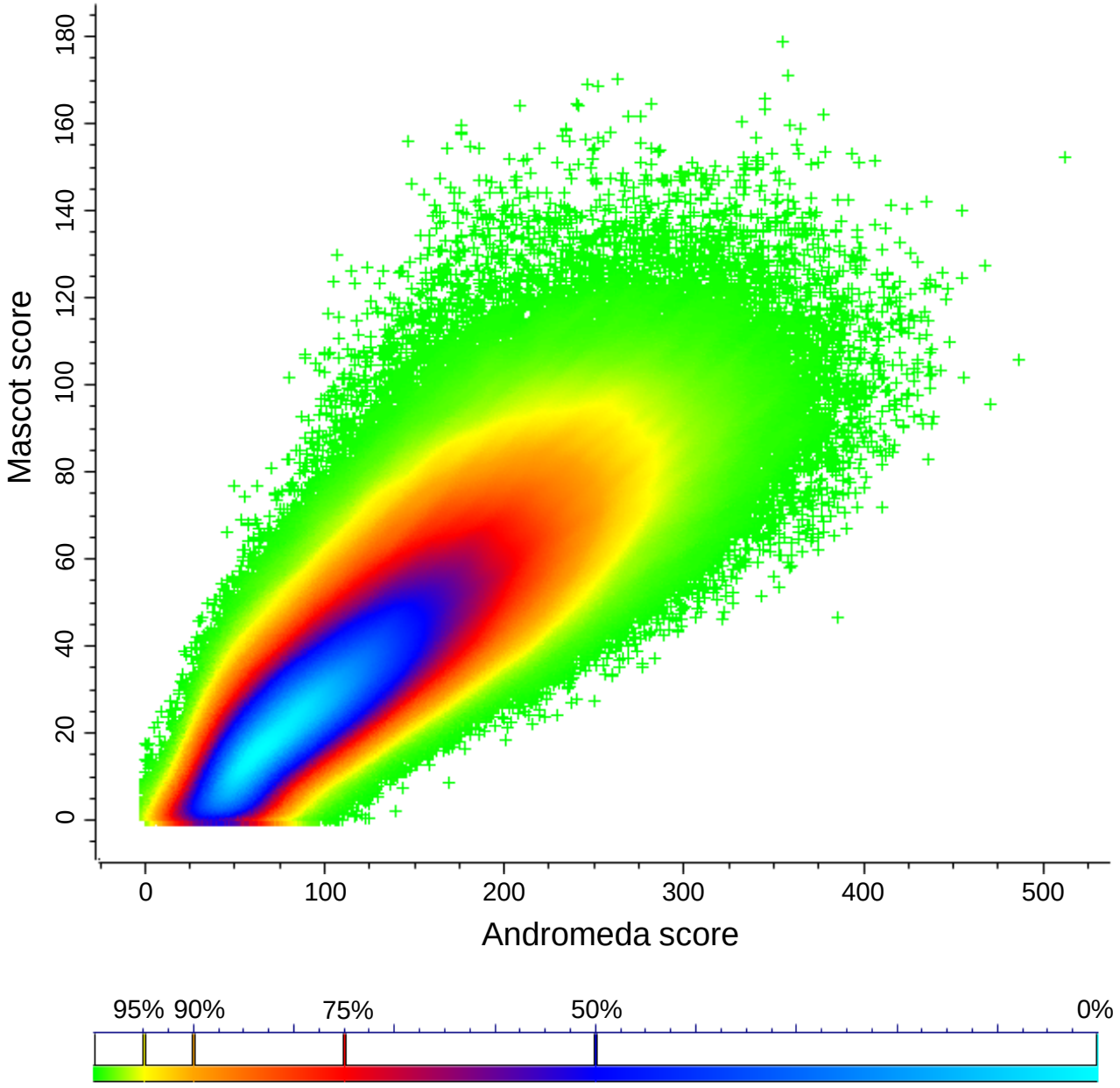
Search engine



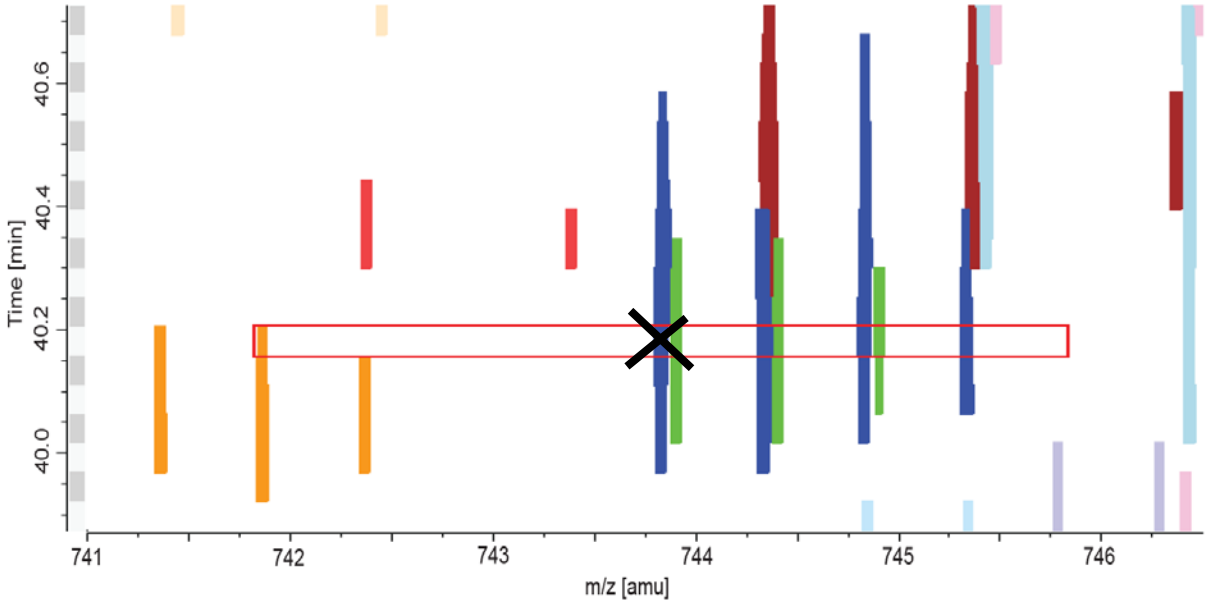


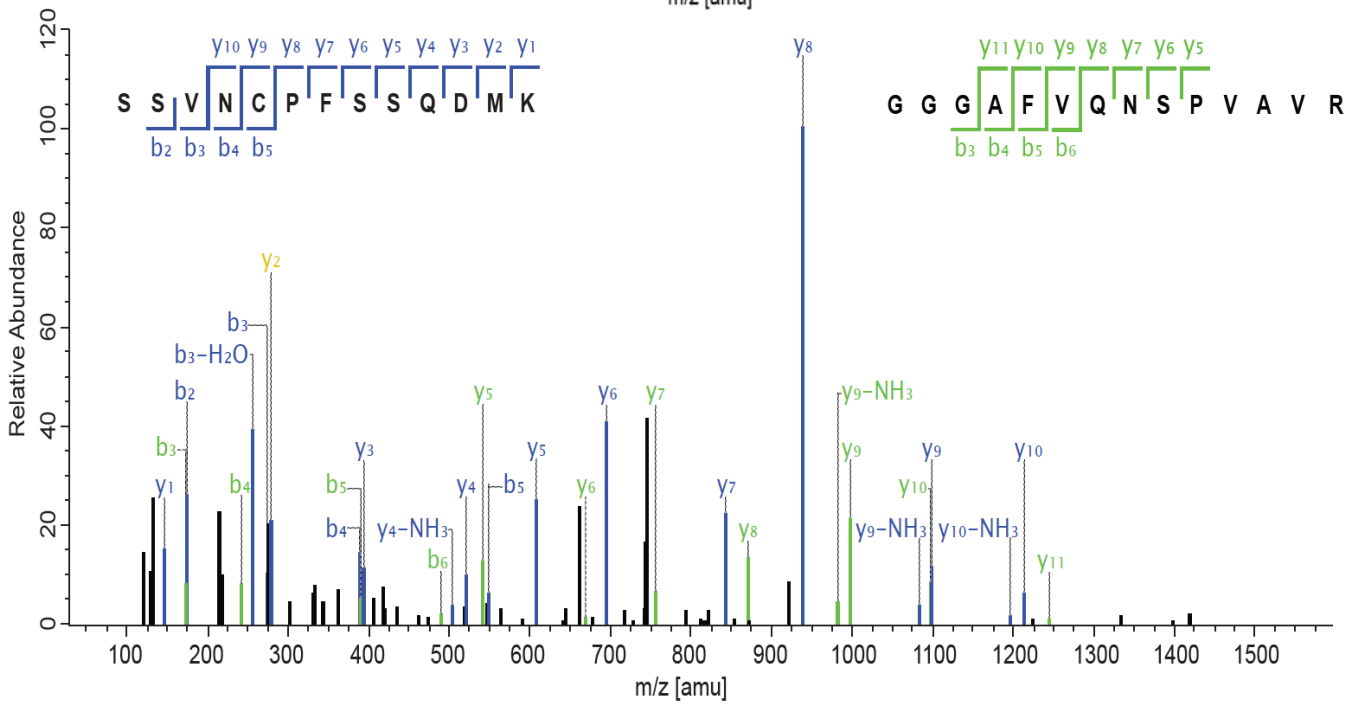
Cox et al, Andromeda – a peptide search engine integrated into the MaxQuant environment. JPR (2011)

Mascot vs. Andromeda score



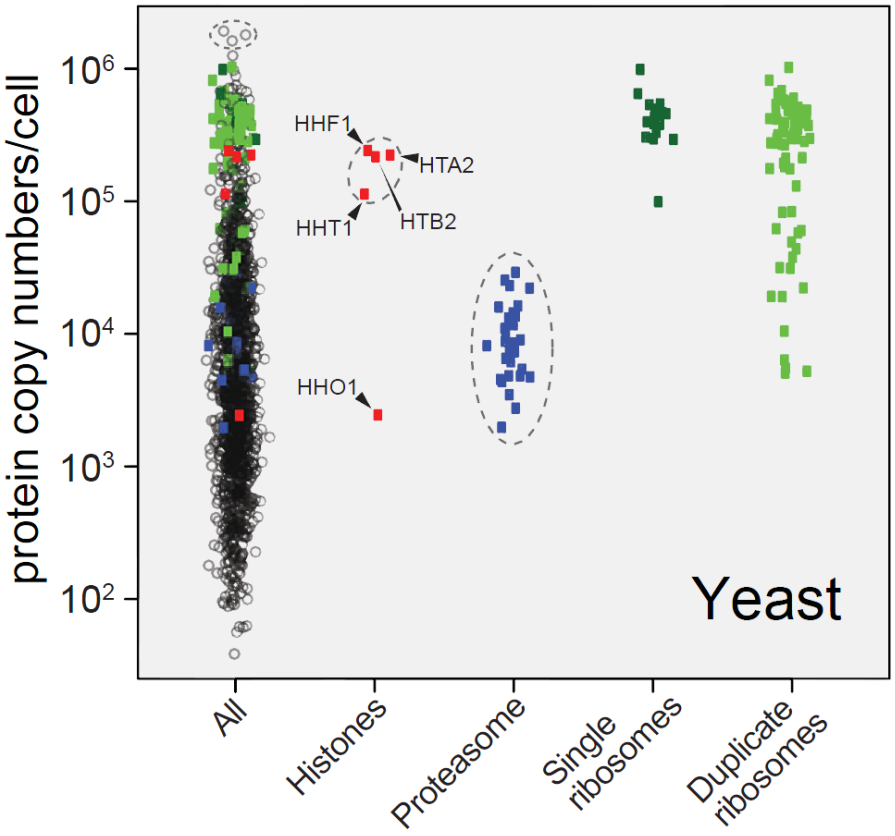
Identification of co-fragmented peptides



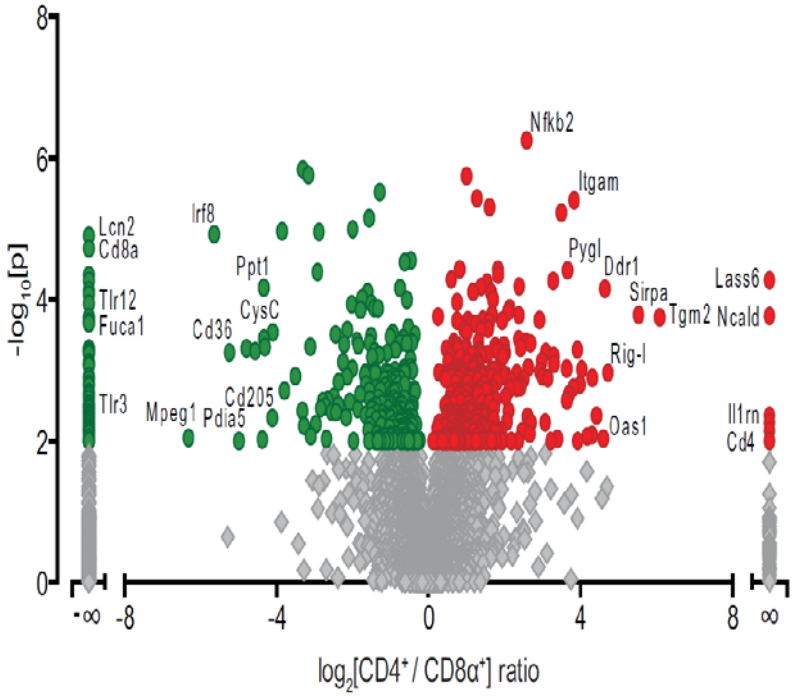


Absolute vs. relative quantification

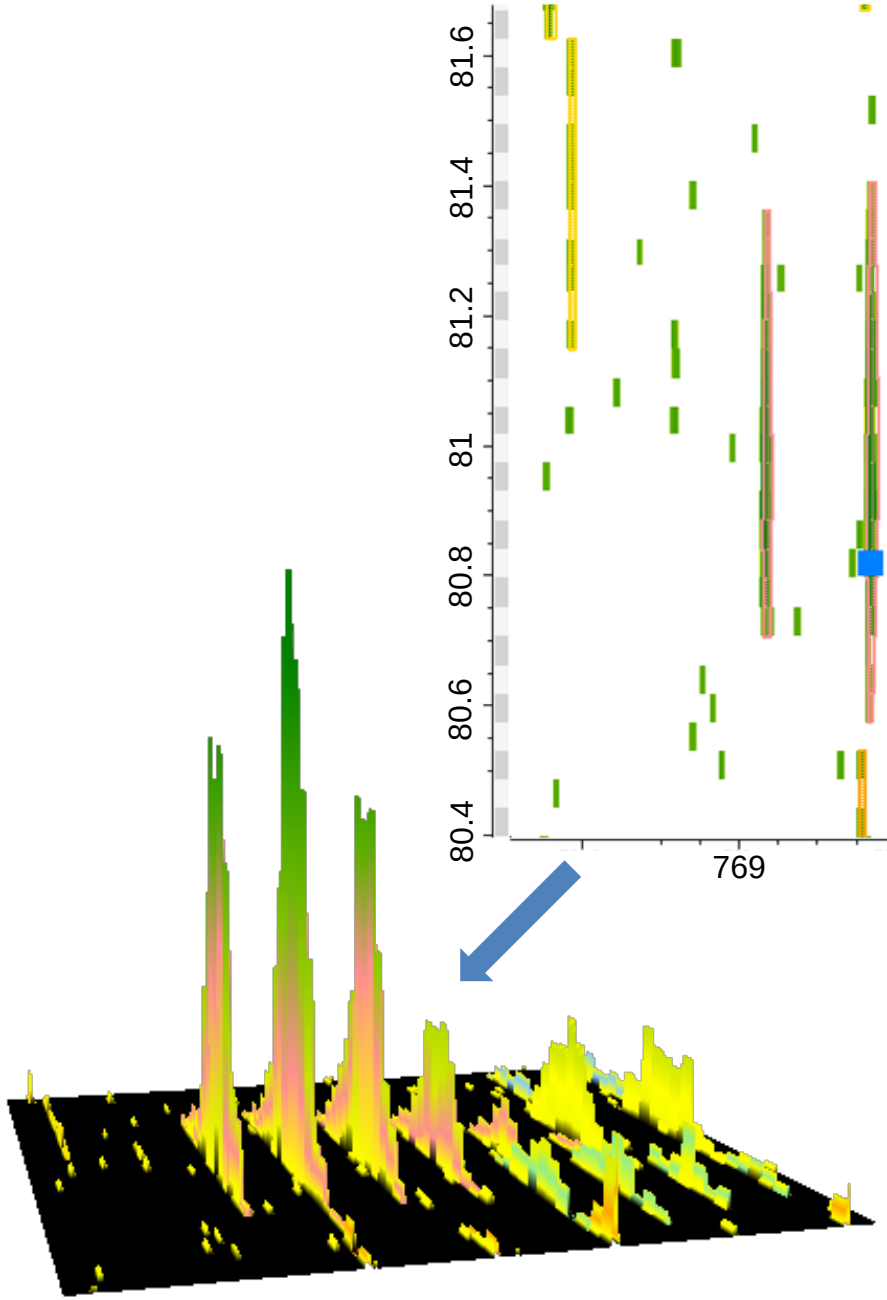
- Absolute quantification: copy numbers for each protein

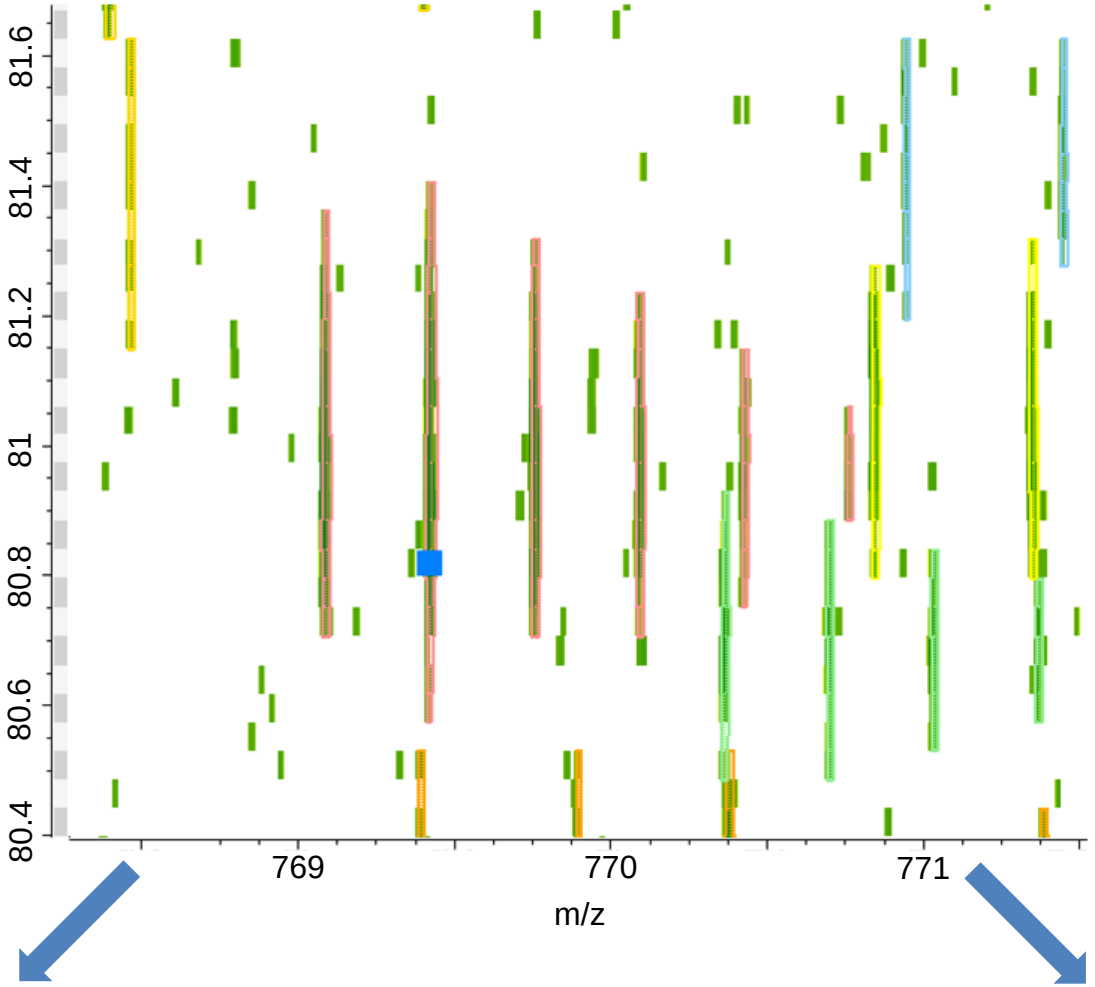


- Relative quantification: compare same protein in different sample



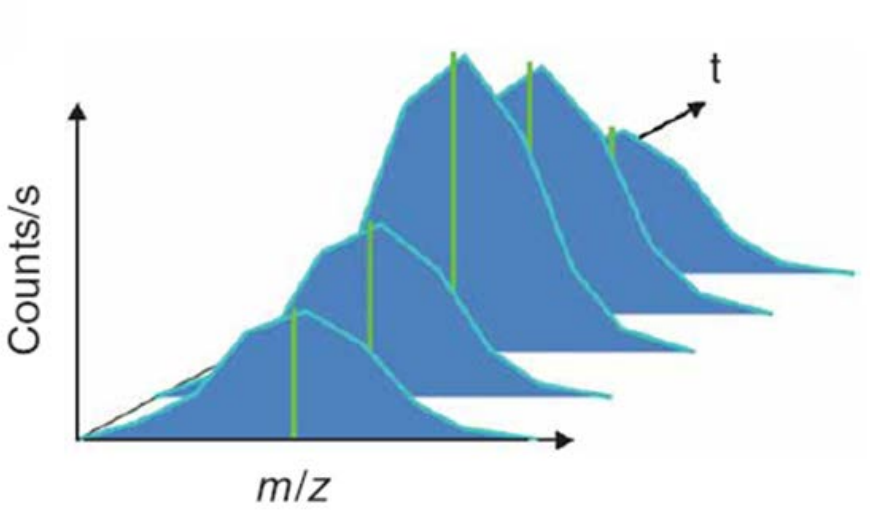
XIC vs. spectral count

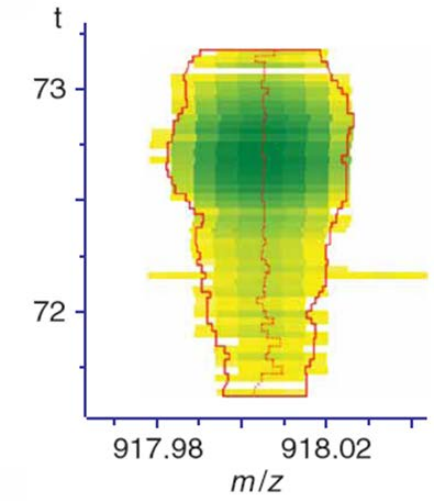


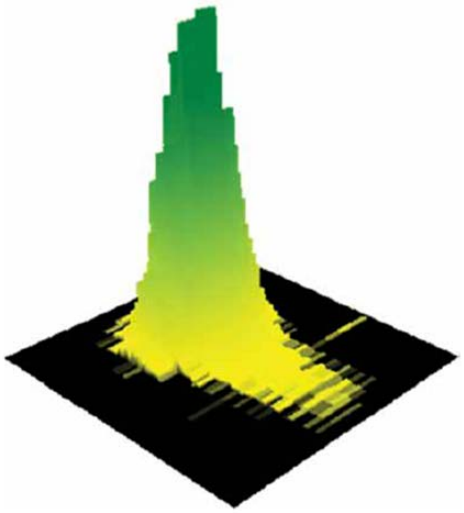


1

3D peak detection

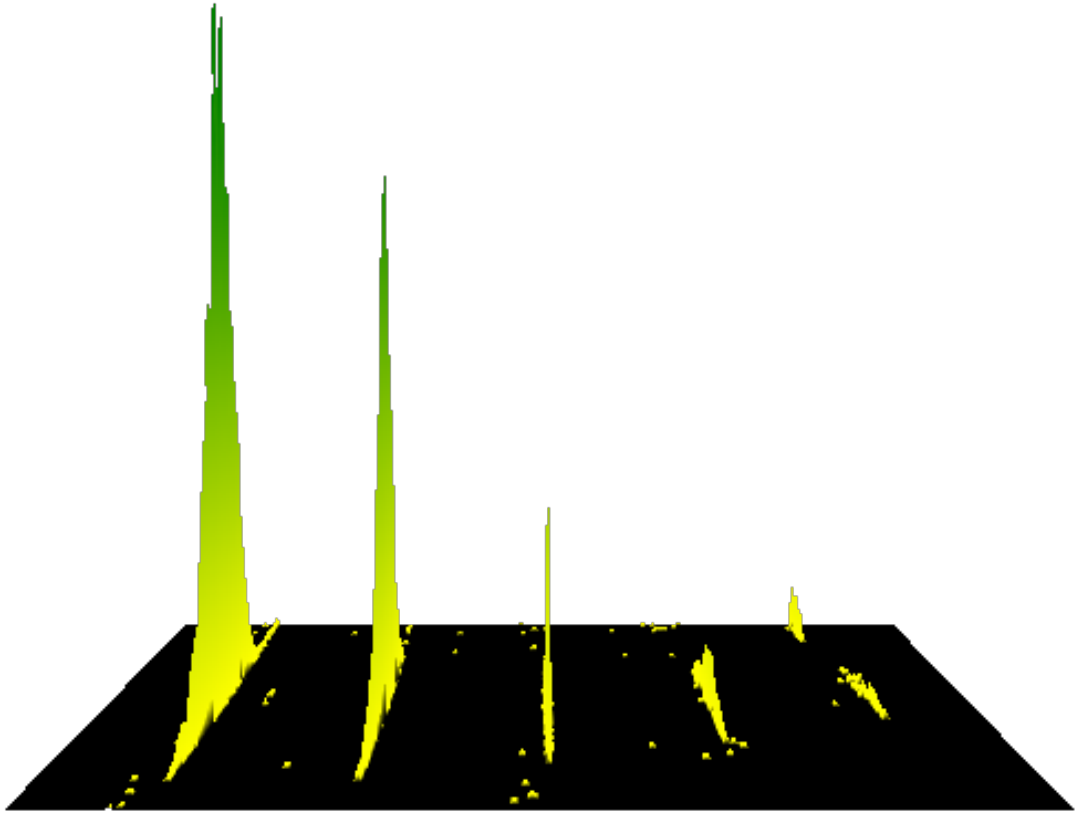






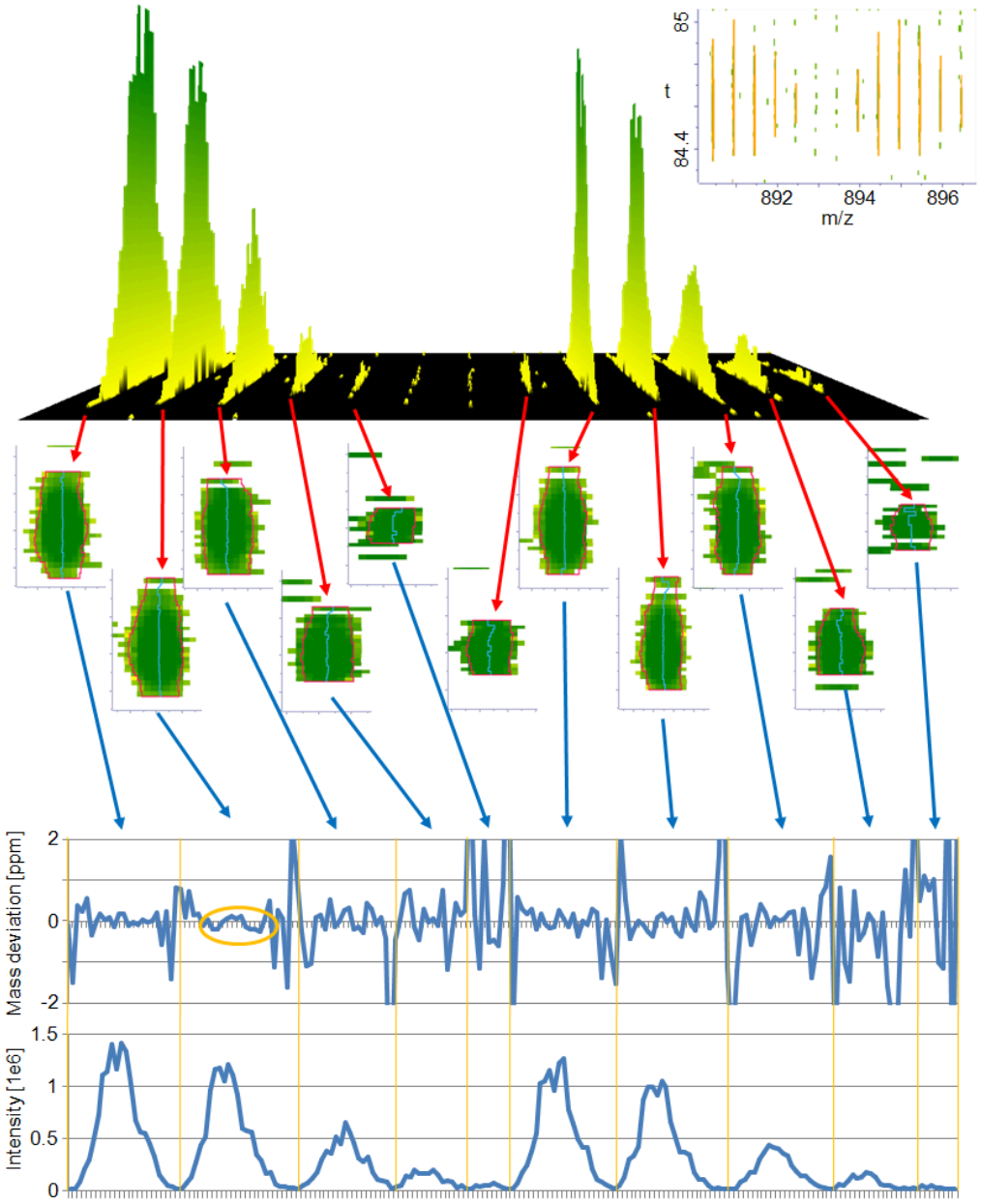
- 2D peaks are assembled into 3D peaks
- Two 2D peaks in adjacent scans are connected when $\Delta m < 7ppm$
- Also next to nearest scan is checked

De-isotoping

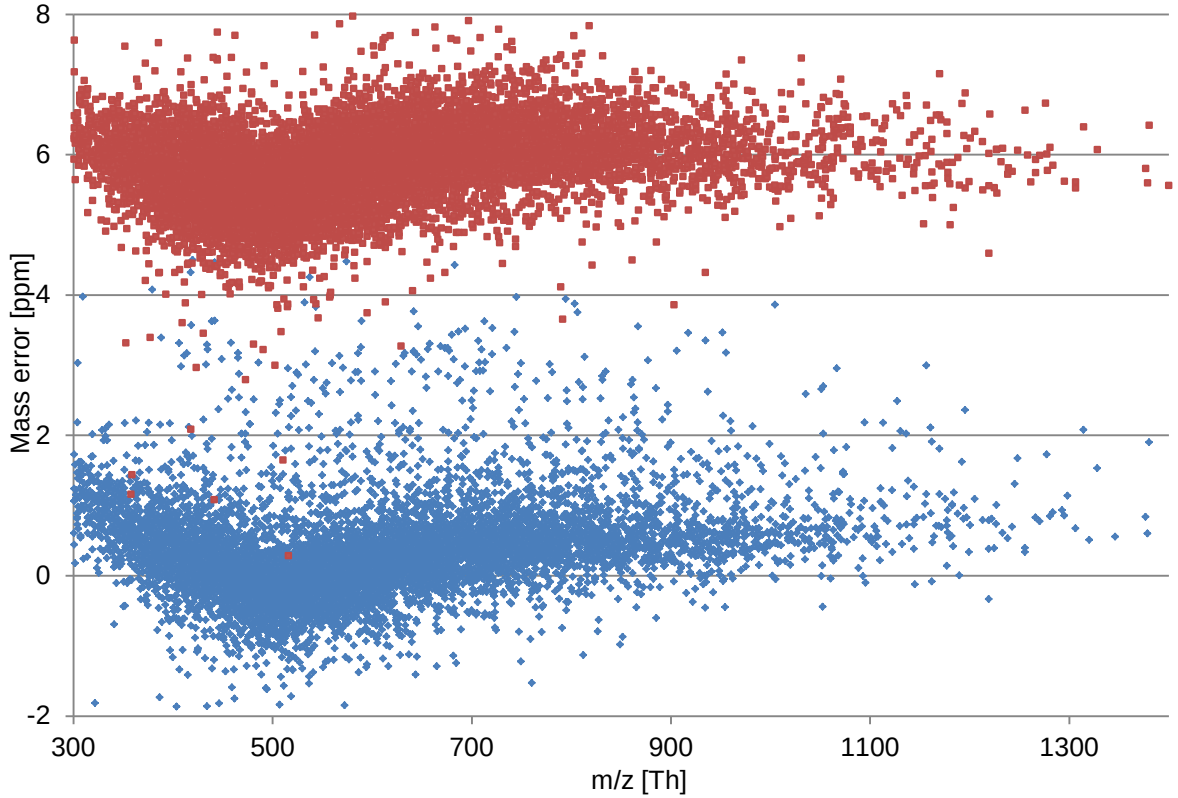


Calculation of precise peptide masses

Calculate precise mean and standard deviation for each peptide mass

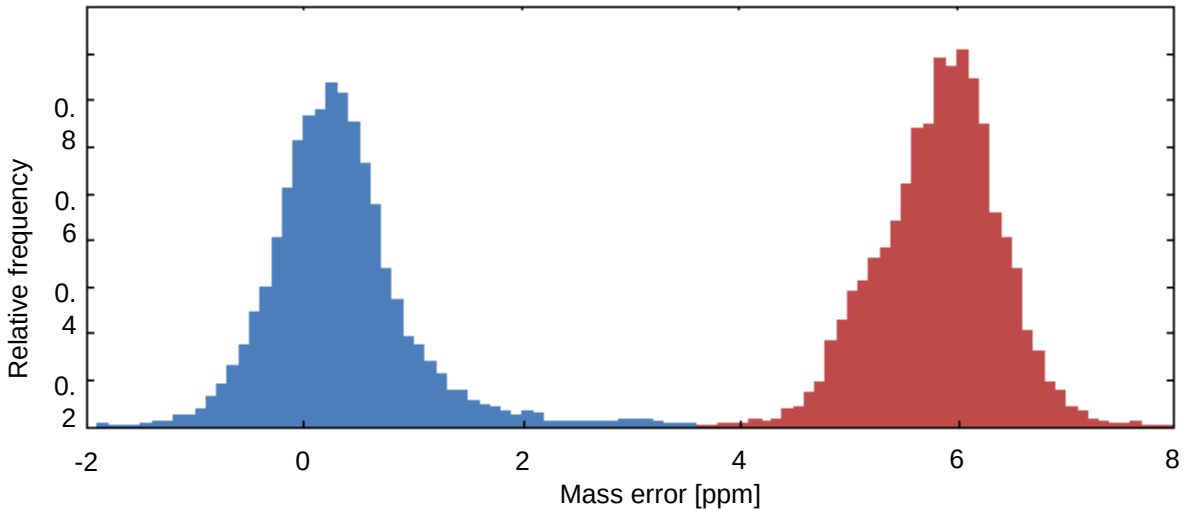


Nonlinear mass recalibration

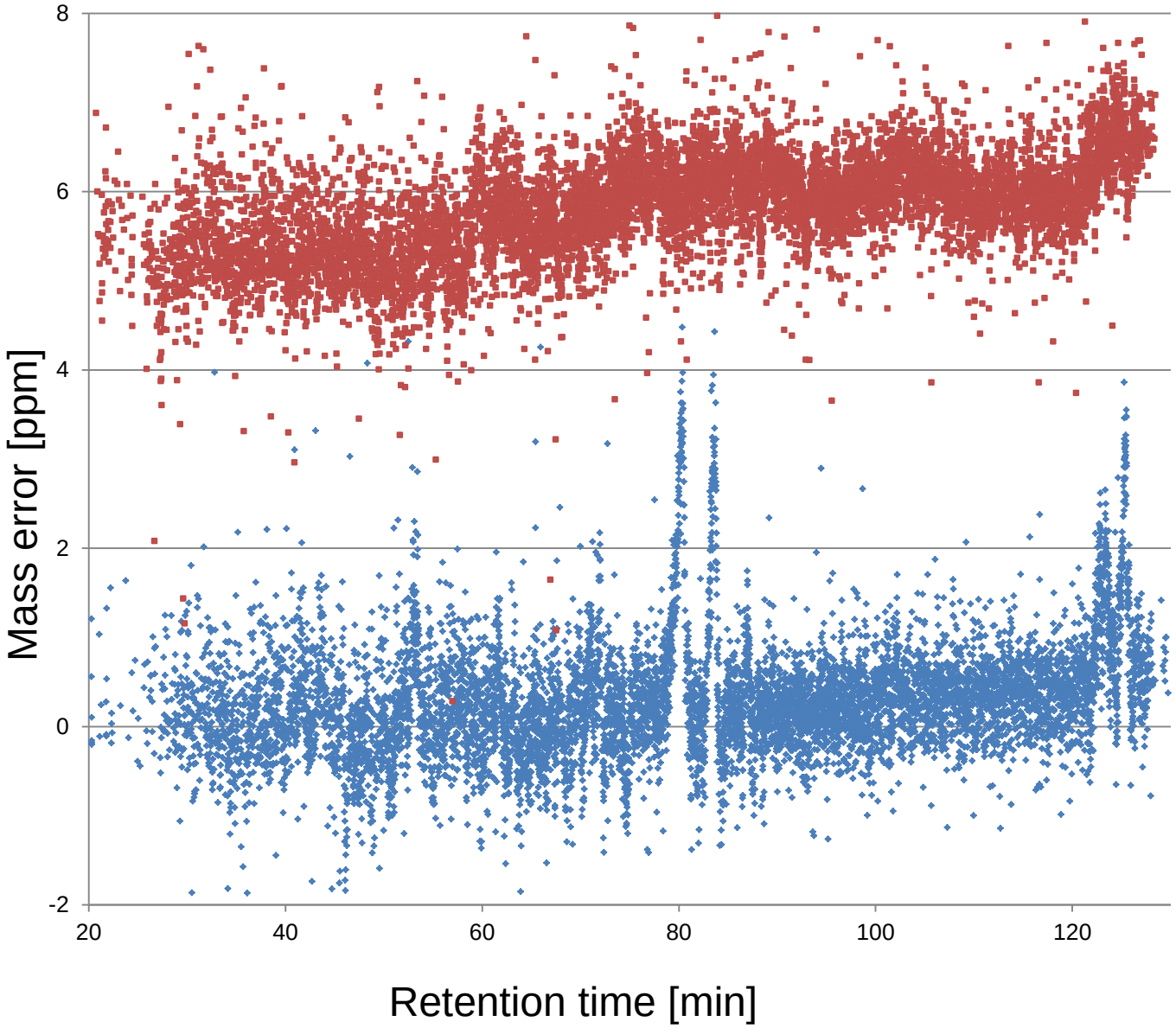


Without lock mass

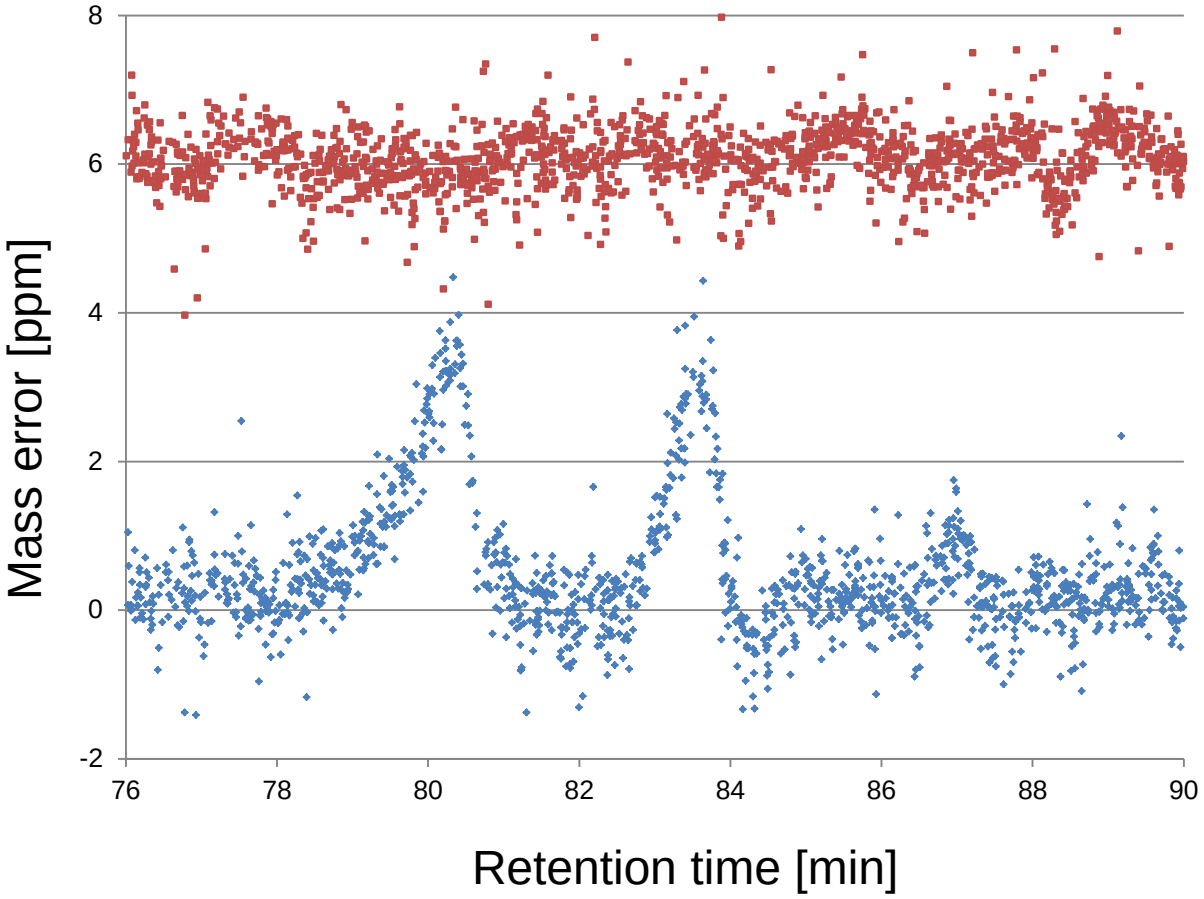
With lock mass



Nonlinear mass recalibration



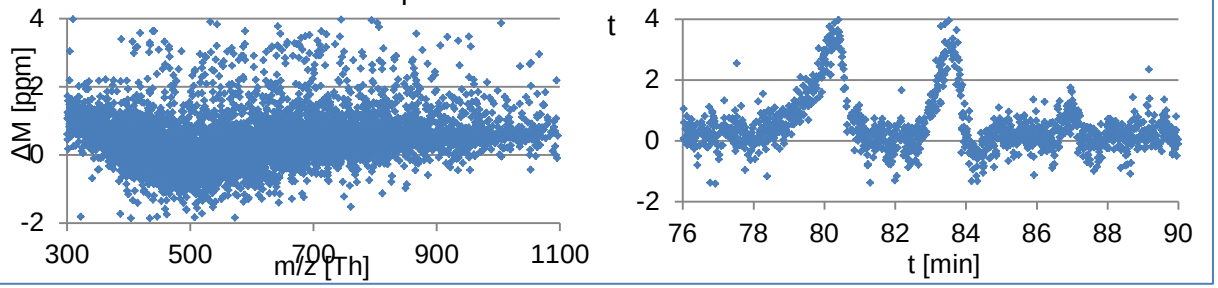
Nonlinear mass recalibration



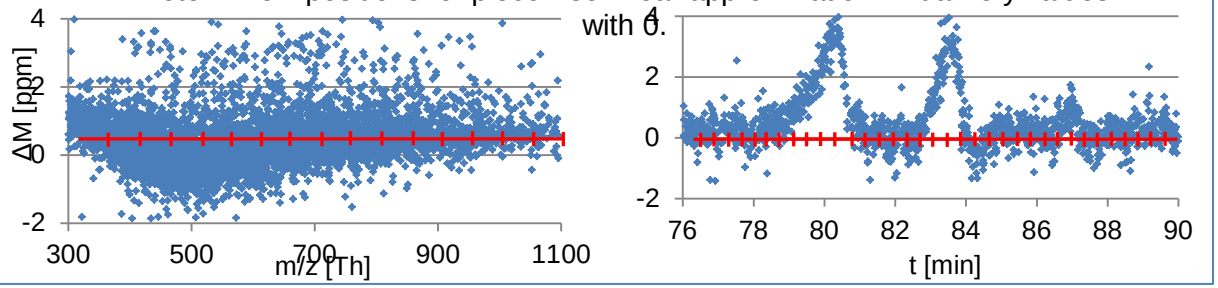
Nonlinear mass recalibration

First Andromeda search with 20ppm mass tolerance and score threshold 80

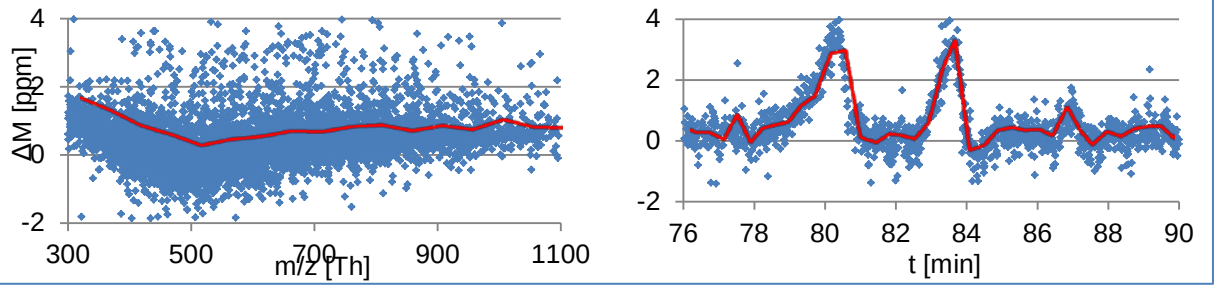
Represent Δm as functions of m/z and t



Determine x positions for piecewise linear approximation. Initialize y values with 0.



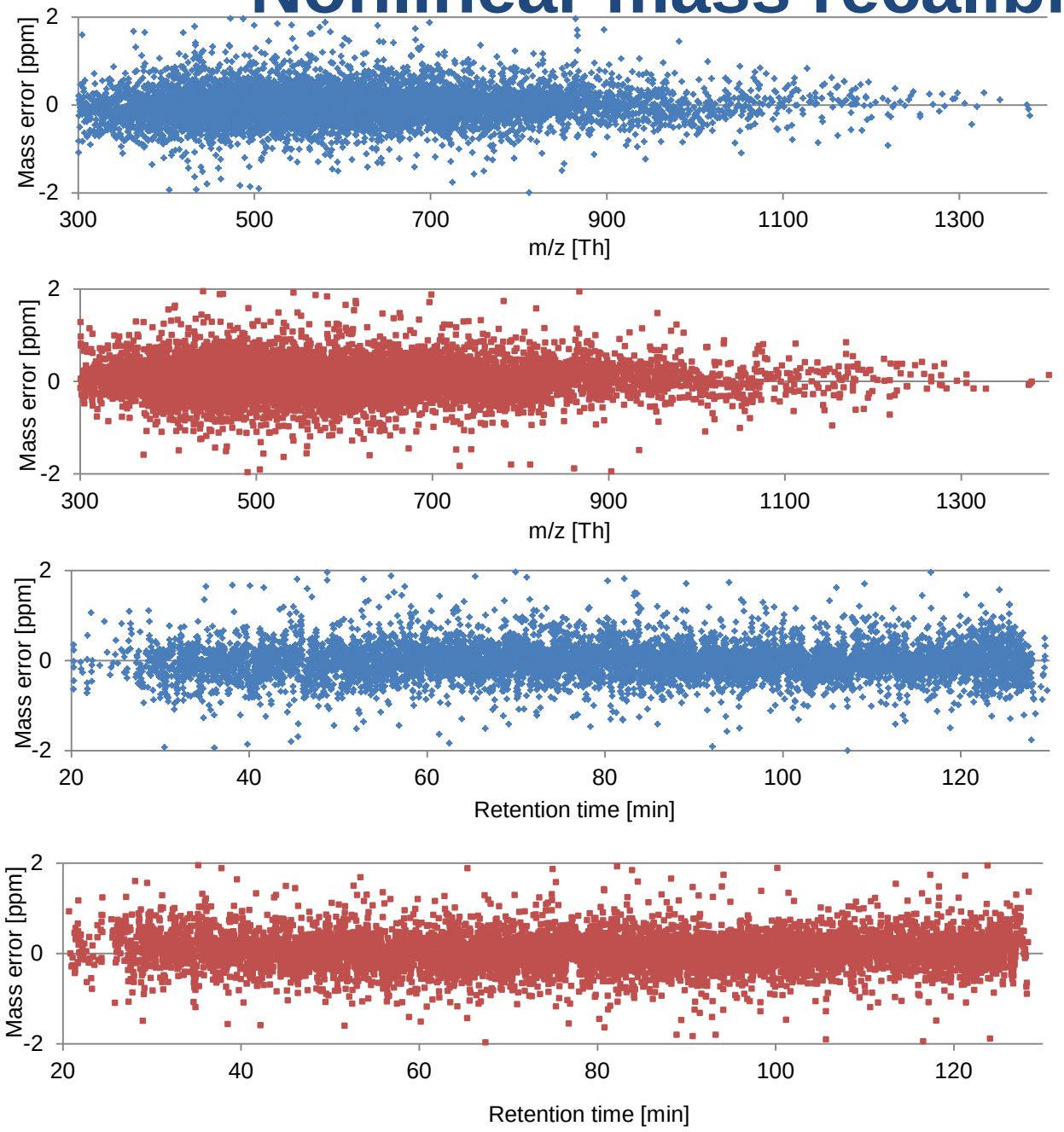
Minimize residual error



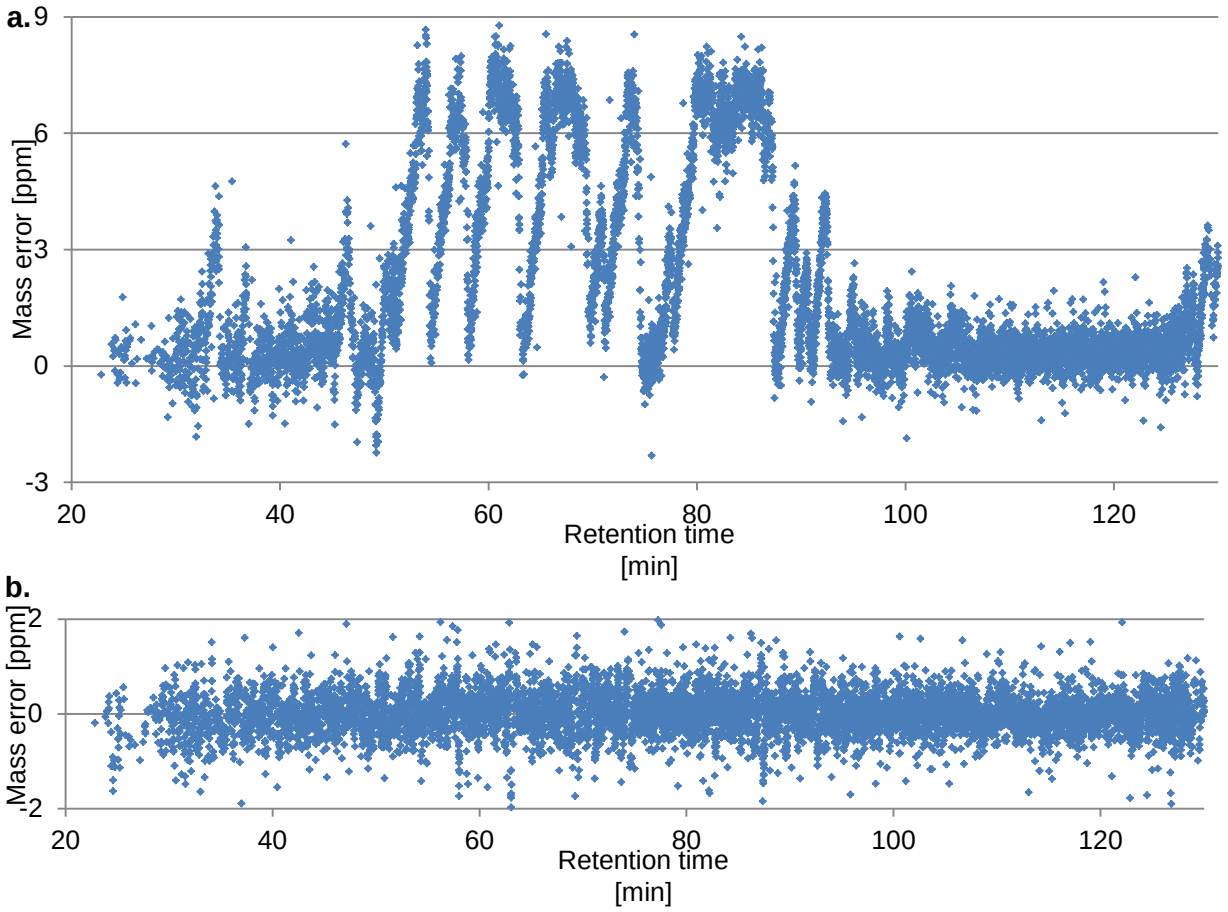
Subtract recalibration functions from all measured peptides

Perform the actual Andromeda search with small individualized mass tolerances

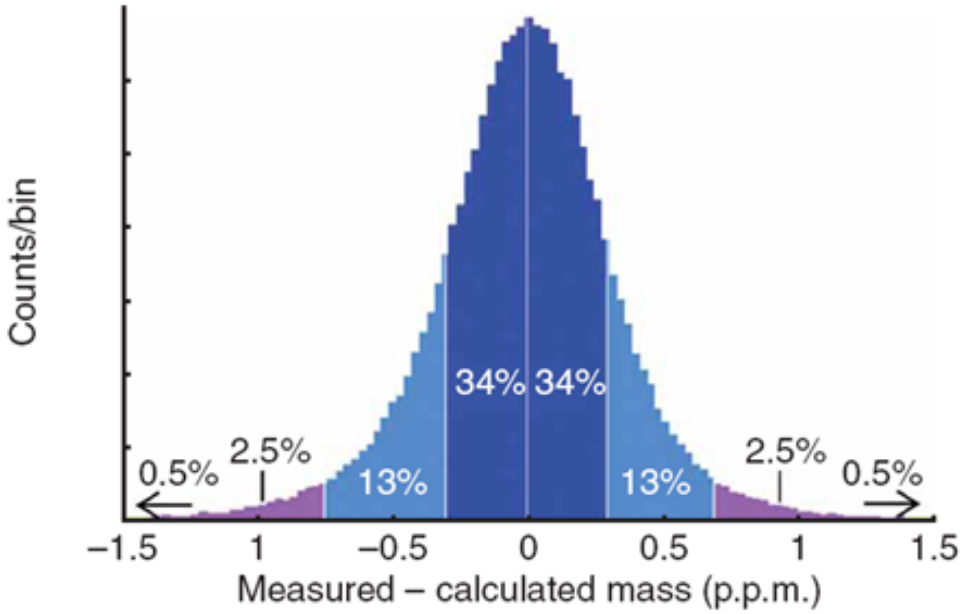
Nonlinear mass recalibration

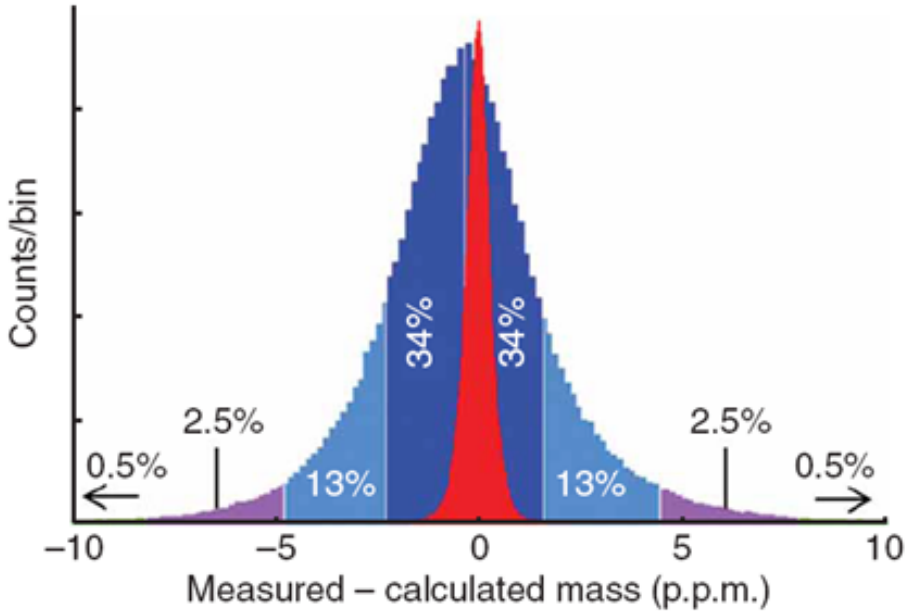


Nonlinear mass recalibration



Improvement in mass accuracy



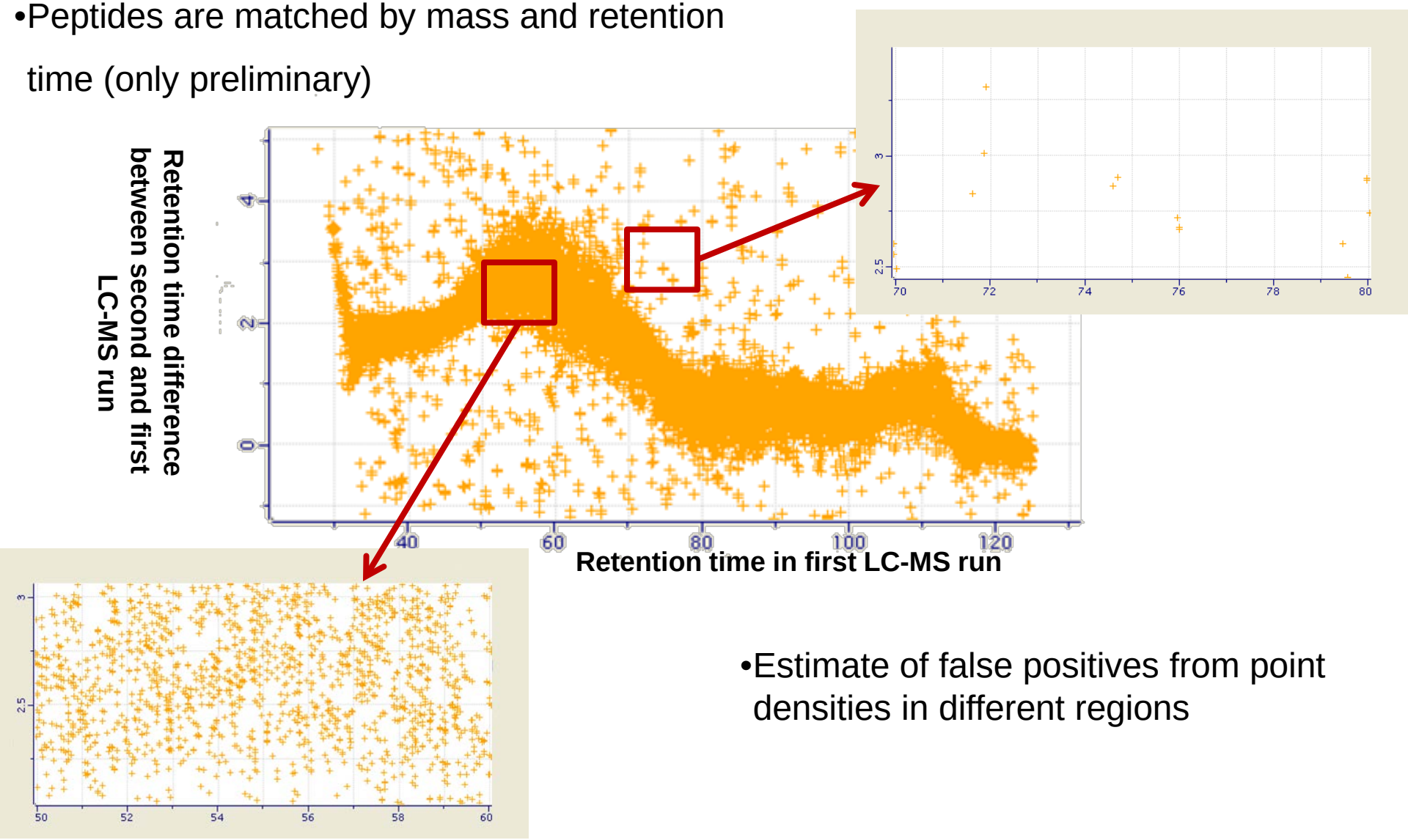


Problems in label free quantification

- Irreproducibility of retention time
- Incompatibility with pre-fractionation
- Quantification in a sample relies on MS/MS identification
- Identified peptides can be different in different samples

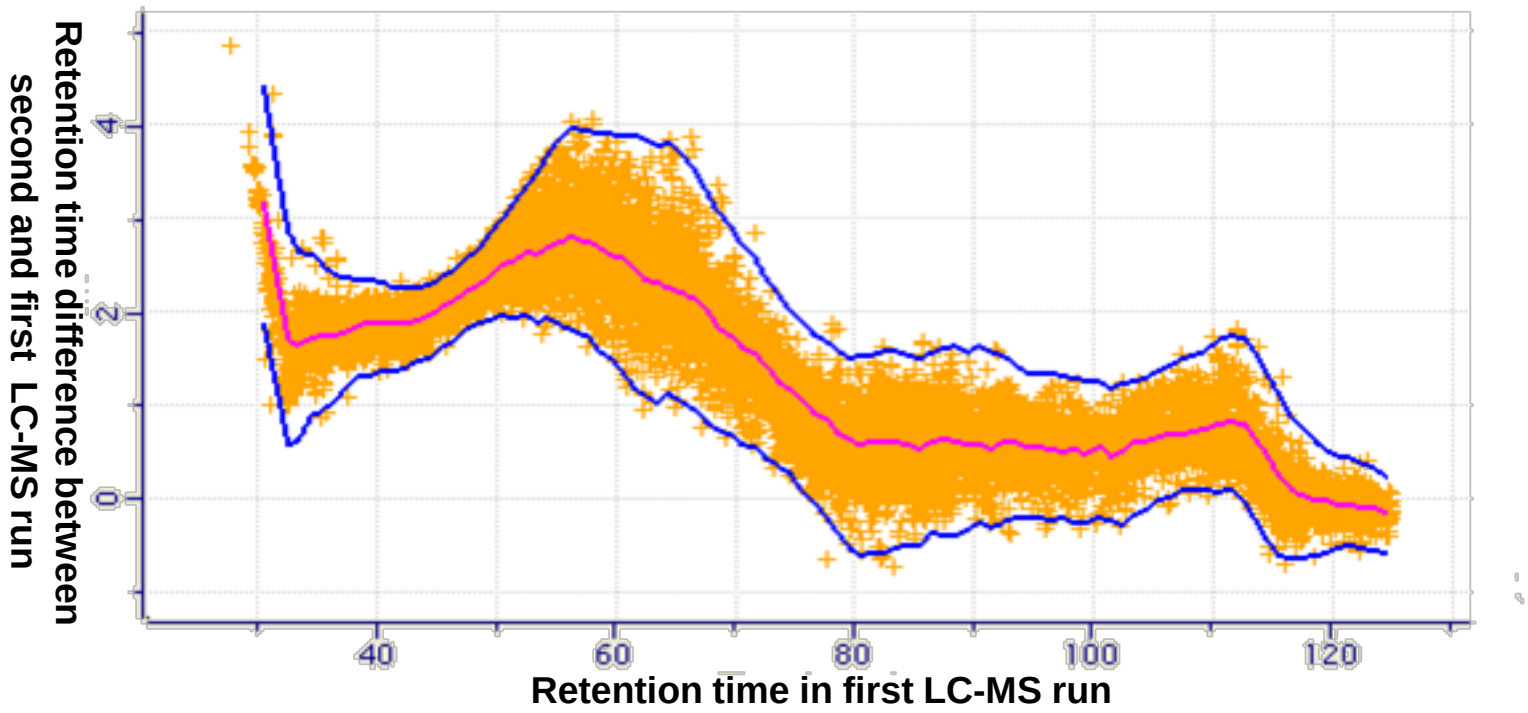
Retention time alignment

- Two LC-MS runs
- Peptides are matched by mass and retention time (only preliminary)

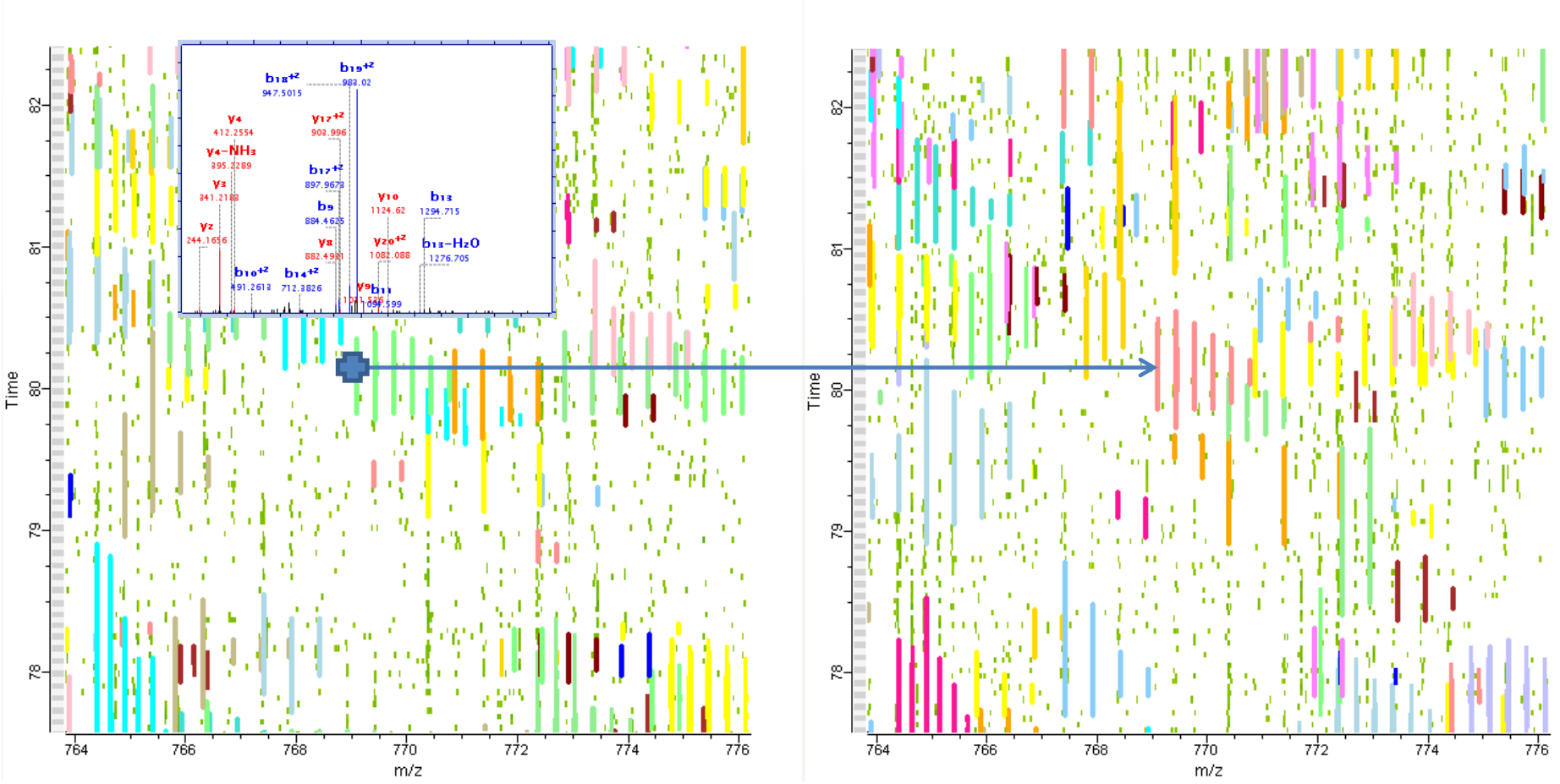


- Estimate of false positives from point densities in different regions

Retention time alignment

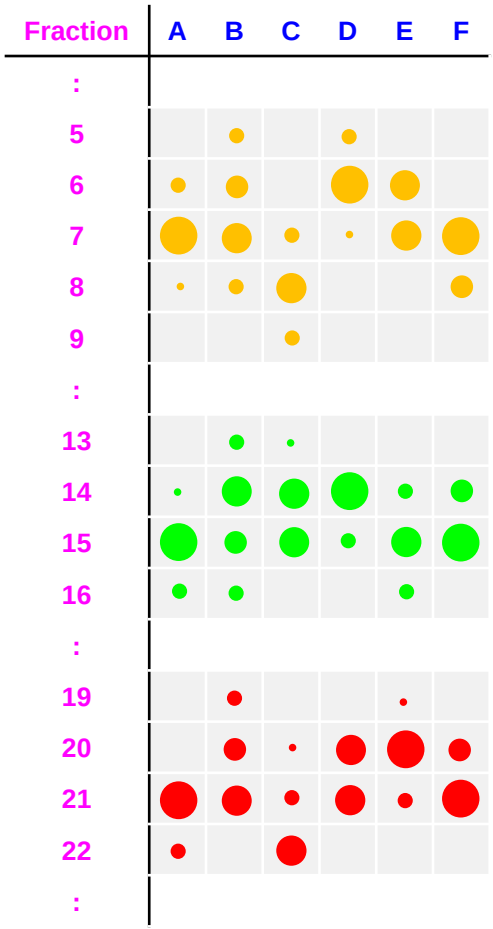


Matching between runs



- Identifcation transfer only between same or adjacend slices/fractions
- Transfering identifications after alignment increases base for quantitation by >100%

Label-free quantification: normalization



Peptide *P*:

$$I_{P,A}(N) = N_{A,6}\,XIC_{A,6} + N_{A,7}\,XIC_{A,7} + N_{A,8}\,XIC_{A,8}$$
$$I_{P,B}(N) = N_{B,5}\,XIC_{B,5} + N_{B,6}\,XIC_{B,6} + N_{B,7}\,XIC_{B,7} + N_{B,8}\,XIC_{B,8}$$
$$I_{P,C}(N) = N_{C,7}\,XIC_{C,7} + N_{C,8}\,XIC_{C,8} + N_{C,9}\,XIC_{C,9}$$
$$I_{P,D}(N) = N_{D,5}\,XIC_{D,5} + N_{D,6}\,XIC_{D,6} + N_{D,7}\,XIC_{D,7}$$
$$I_{P,E}(N) = N_{E,6}\,XIC_{E,6} + N_{E,7}\,XIC_{E,7}$$
$$I_{P,F}(N) = N_{F,7}\,XIC_{F,7} + N_{F,8}\,XIC_{F,8}$$

Peptide *Q*:

$$I_{Q,A}(N) = N_{A,14}\,XIC_{A,14} + N_{A,15}\,XIC_{A,15} + N_{A,16}\,XIC_{A,16}$$
$$I_{Q,B}(N) = N_{B,13}\,XIC_{B,13} + N_{B,14}\,XIC_{B,14} + N_{B,15}\,XIC_{B,15} + N_{B,16}\,XIC_{B,16}$$
$$I_{Q,C}(N) = N_{C,13}\,XIC_{C,13} + N_{C,14}\,XIC_{C,14} + N_{C,15}\,XIC_{C,15}$$
$$I_{Q,D}(N) = N_{D,14}\,XIC_{D,14} + N_{D,15}\,XIC_{D,15}$$
$$I_{Q,E}(N) = N_{E,14}\,XIC_{E,14} + N_{E,15}\,XIC_{E,15} + N_{E,16}\,XIC_{E,16}$$
$$I_{Q,F}(N) = N_{F,14}\,XIC_{F,14} + N_{F,15}\,XIC_{F,15}$$

Peptide *R*:

$$I_{R,A}(N) = N_{A,21}\,XIC_{A,21} + N_{A,22}\,XIC_{A,22}$$
$$I_{R,B}(N) = N_{B,19}\,XIC_{B,19} + N_{B,20}\,XIC_{B,20} + N_{B,21}\,XIC_{B,21}$$
$$I_{R,C}(N) = N_{C,20}\,XIC_{C,20} + N_{C,21}\,XIC_{C,21} + N_{C,22}\,XIC_{C,22}$$
$$I_{R,D}(N) = N_{D,20}\,XIC_{D,20} + N_{D,21}\,XIC_{D,21}$$
$$I_{R,E}(N) = N_{E,19}\,XIC_{E,19} + N_{E,20}\,XIC_{E,20} + N_{E,21}\,XIC_{E,21}$$
$$I_{R,F}(N) = N_{F,20}\,XIC_{F,20} + N_{F,21}\,XIC_{F,21}$$

$$H_P(N) = \left|\log\left(\frac{I_{P,A}(N)}{I_{P,B}(N)}\right)\right|^2 + \left|\log\left(\frac{I_{P,A}(N)}{I_{P,C}(N)}\right)\right|^2 + \left|\log\left(\frac{I_{P,A}(N)}{I_{P,D}(N)}\right)\right|^2 + \text{other sample pairs}$$

$$H_Q(N) = \left|\log\left(\frac{I_{Q,A}(N)}{I_{Q,B}(N)}\right)\right|^2 + \left|\log\left(\frac{I_{Q,A}(N)}{I_{Q,C}(N)}\right)\right|^2 + \left|\log\left(\frac{I_{Q,A}(N)}{I_{Q,D}(N)}\right)\right|^2 + \text{other sample pairs}$$

$$H_R(N) = \left|\log\left(\frac{I_{R,A}(N)}{I_{R,B}(N)}\right)\right|^2 + \left|\log\left(\frac{I_{R,A}(N)}{I_{R,C}(N)}\right)\right|^2 + \left|\log\left(\frac{I_{R,A}(N)}{I_{R,D}(N)}\right)\right|^2 + \text{other sample pairs}$$

$$H(N) = H_P(N) + H_Q(N) + H_R(N) + \text{other peptides}$$

Protein quantification

a.

>P63208
MPSIKLQSSDGEIFEVDVEIAKQSVTIKTMLEDLGMKDEGDD
DPVPLPNVNAAILKKVIQWCTHHKDDPPPPEDDENKEKRTDD
IPVWDQEFLKVDQGTLFELILAANYLDIKGLLDVTCKTVANM
IKGKTPEEIRKTFNIKNDFTEEEEAQVRKENQWCEEK

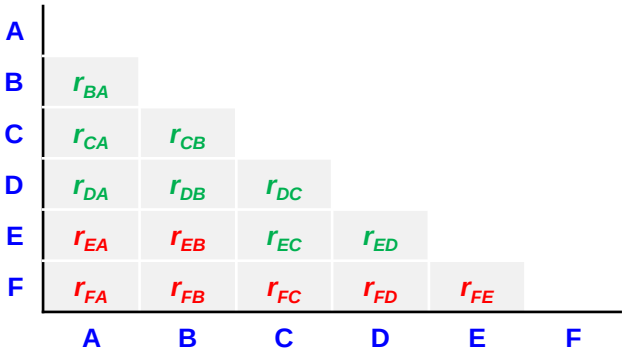
b.

Peptide	Sequence
P_1	LQSSDGEIFEVDVEIAK
P_2	TMLEDLGMK
P_3	VIQWCTHHK
P_4	RTDDIPVWDQEFLK
P_5	TVANMIK
P_6	TPEEIRK
P_7	NDFTEEEEAQVR

c.

Sample	P_1	P_2	P_3	P_4	P_5	P_6	P_7
A		+				+	
B		+	+			+	
C	+	+	+	+		+	+
D	+	+		+		+	+
E		+		+			+
F		+			+		

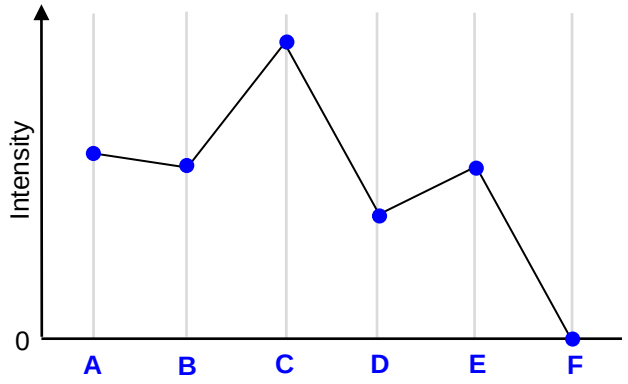
d.



e.

$r_{BA} = I_B / I_A$	$r_{CA} = I_C / I_A$	$r_{CB} = I_C / I_B$
$r_{DA} = I_D / I_A$	$r_{DB} = I_D / I_B$	$r_{DC} = I_D / I_C$
$r_{EC} = I_E / I_C$	$r_{ED} = I_E / I_D$	$I_F = 0$

f.



Label-free quantification
Benchmark dataset

- HeLa and E. coli cell lysates are mixed
- Proteins were digested with trypsin.
- In three replicates peptides were separated by isoelectric focusing in 24 fractions.
- This was repeated with the same amount of HeLa, but E. coli lysate tripled.
- This results in six samples for which all human proteins have constant protein profiles, while E. coli proteins have a ratio of three between replicate groups.
- LC-MS on an LTQ-Orbitrap mass spectrometer.

Data: Christian Luber

Identification results

1,234,125 MS isotope patterns identified by MS/MS

1,852,556 MS isotope patterns identified by matching between runs

3,086,681 MS isotope patterns in total

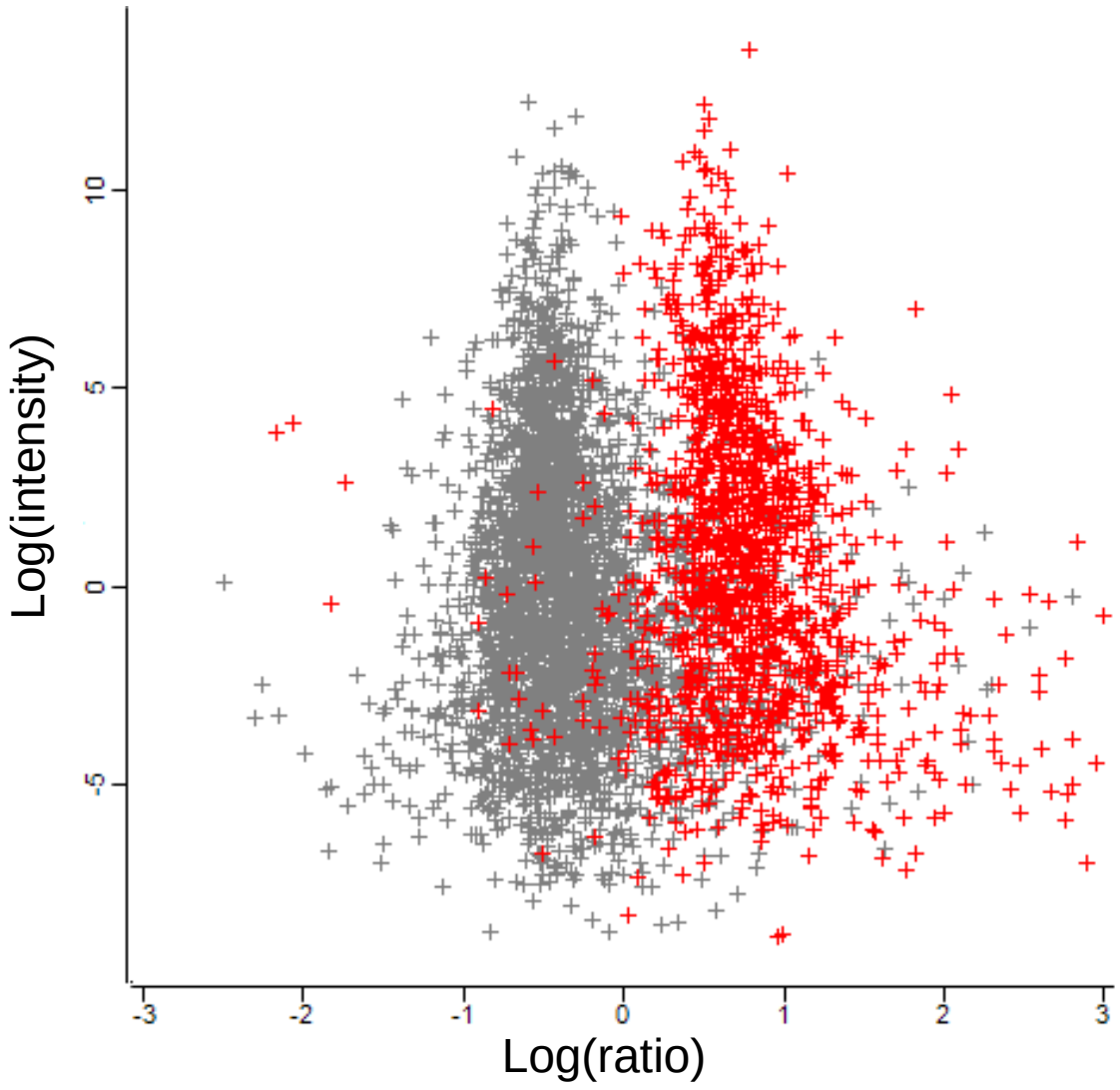
6,577 proteins

5,161 proteins in at least 3/6 samples

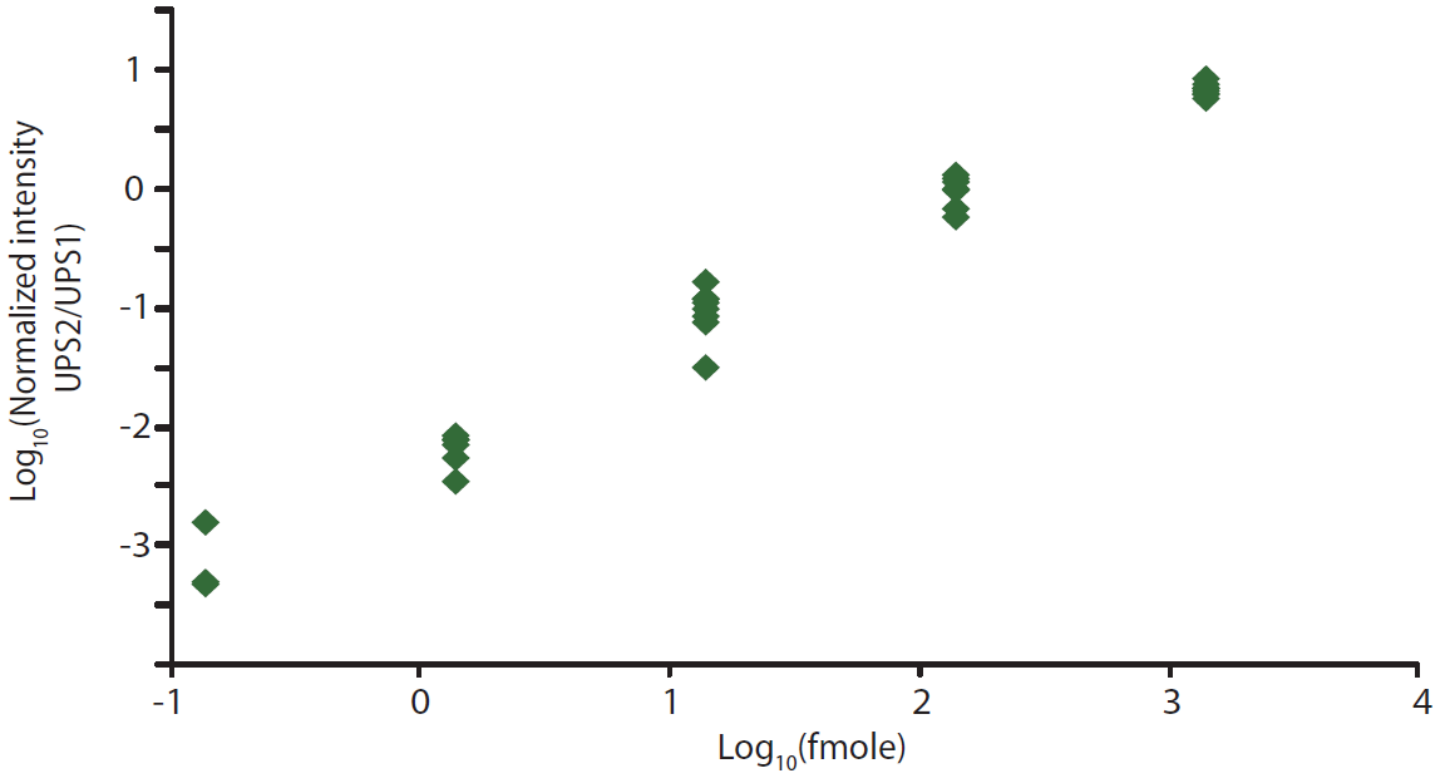
4,589 proteins in 6/6 samples

46,839 peptide sequences

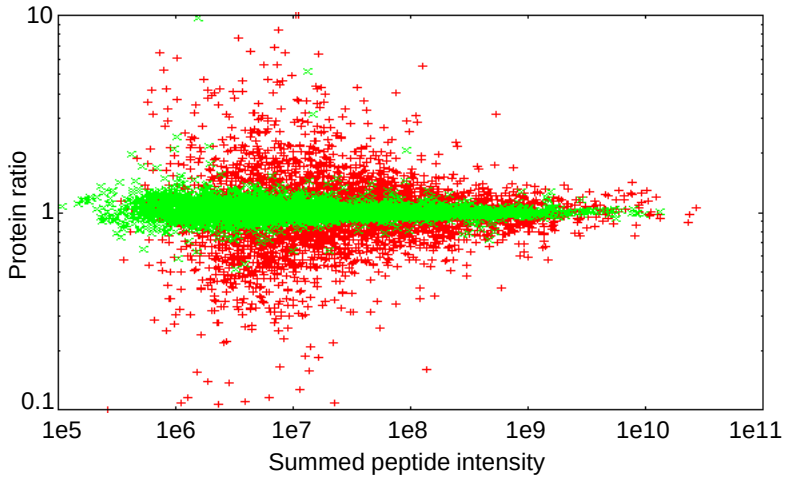
Label-free quantification results



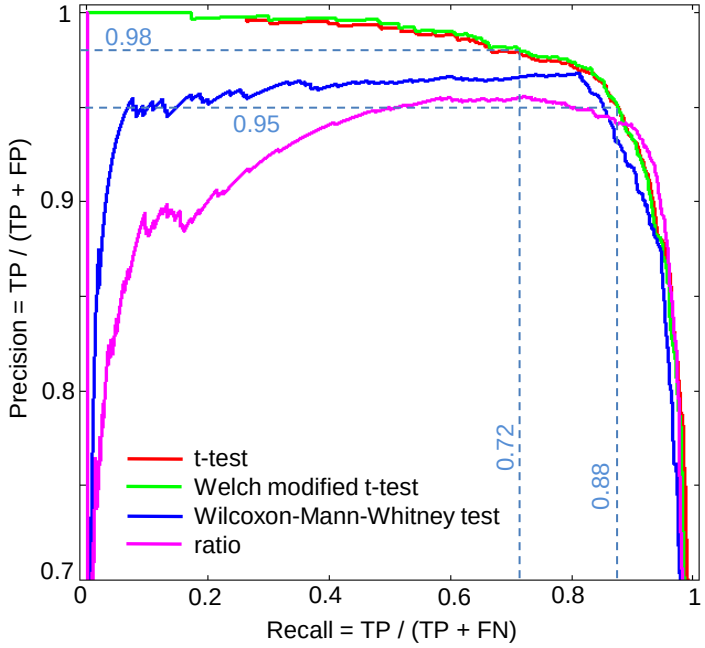
Dynamic range benchmark dataset



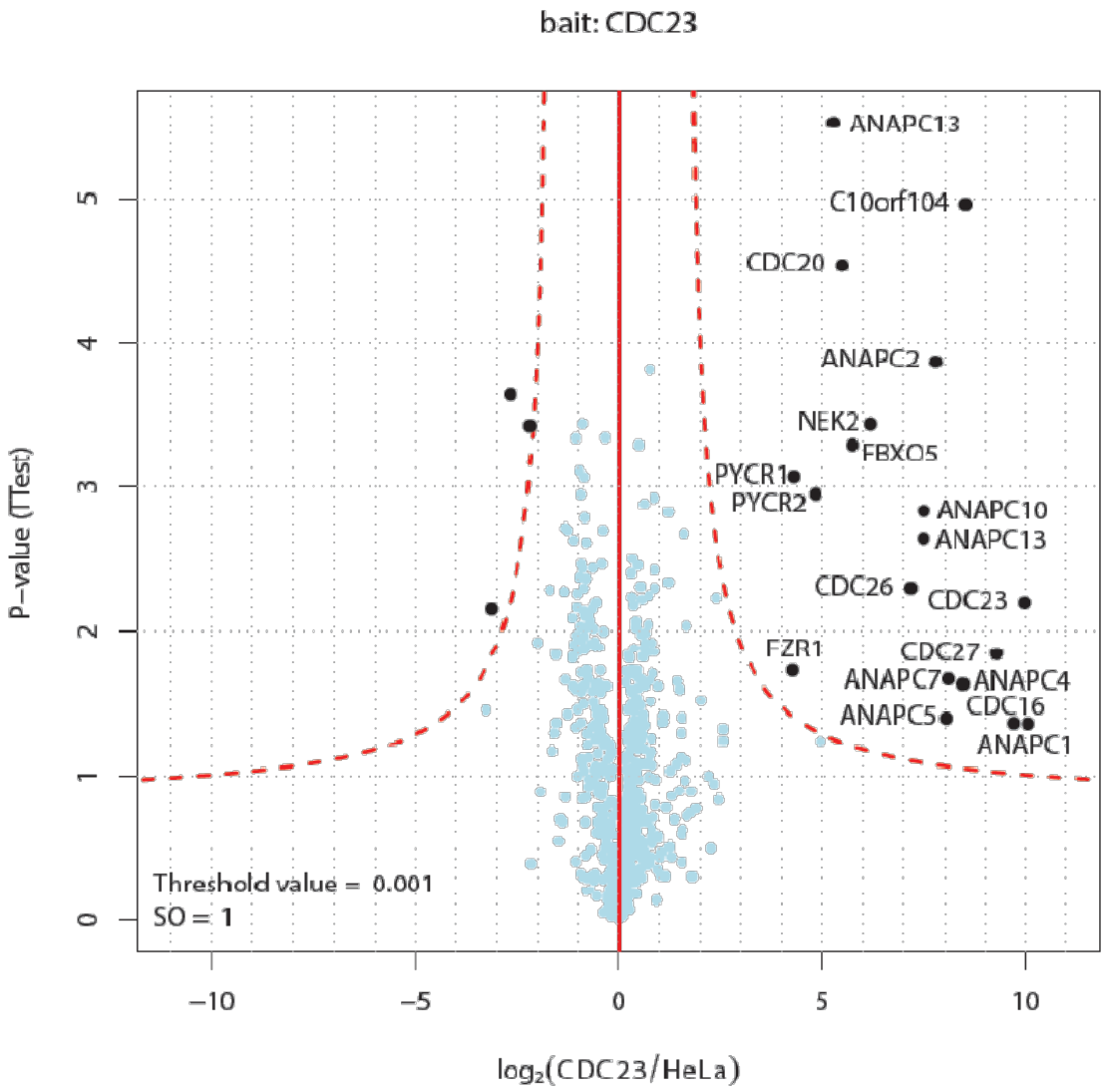
Comparison to SILAC



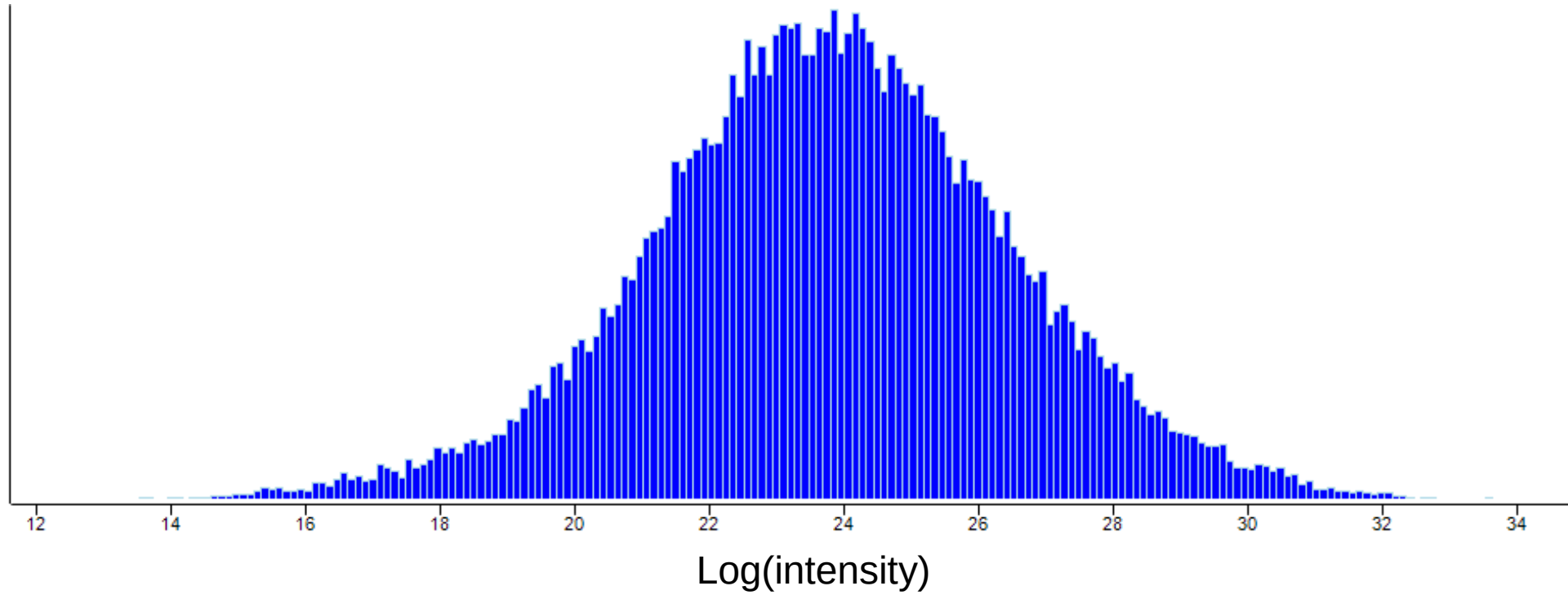
Precision vs. recall



Pulldowns

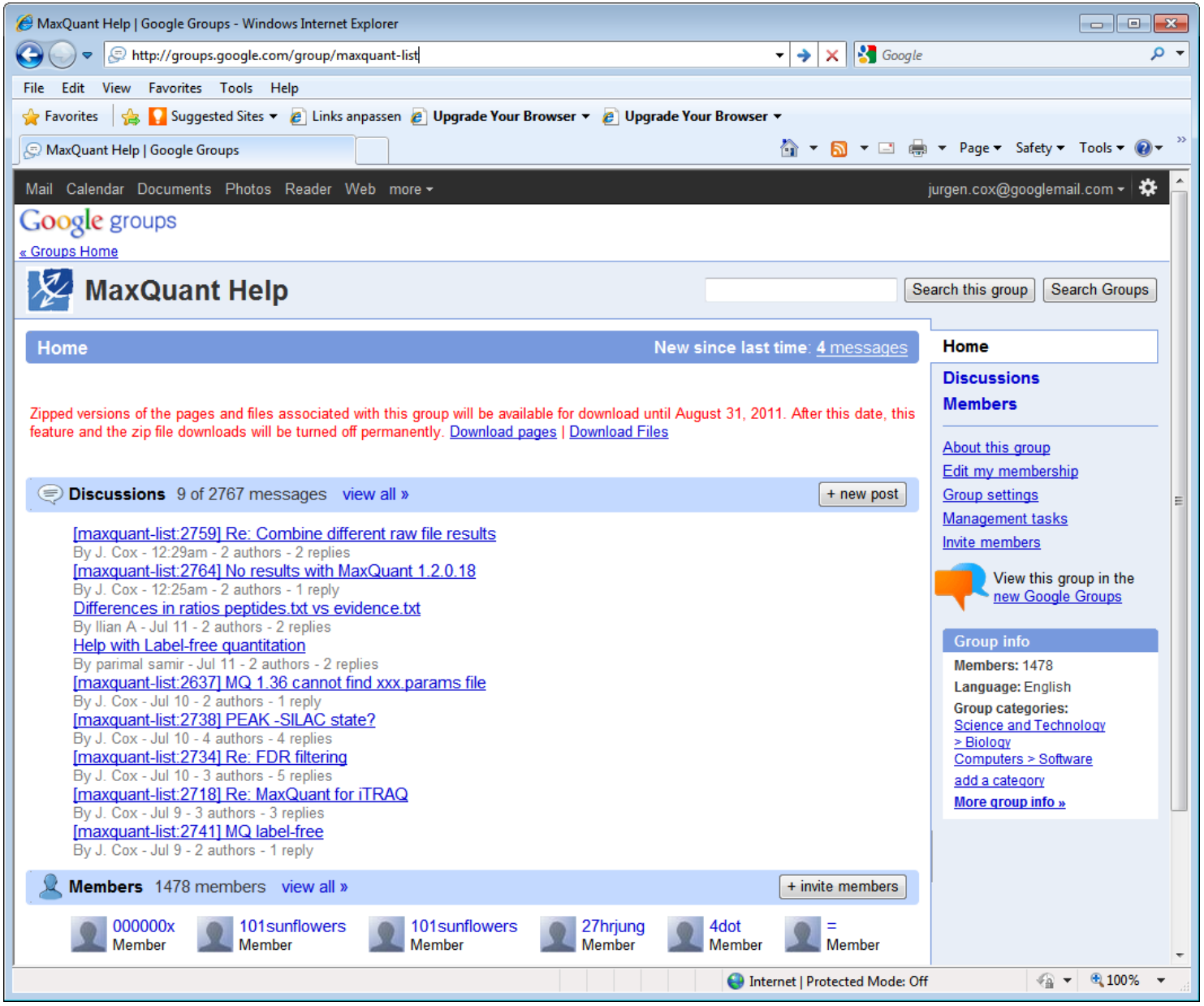

bait: CDC23
ANAPC13
C10orf104
CDC20
ANAPC2
NEK2
FBXO5
PYCR1
PYCR2
ANAPC10
ANAPC13
CDC26
CDC23
FZR1
CDC27
ANAPC7
ANAPC4
ANAPC5
CDC16
ANAPC1
Threshold value = 0.001
SO = 1
P−value (TTest)
0
1
2
3
4
5
−10
−5
0
5
10
log2(CDC23/HeLa)

Imputation

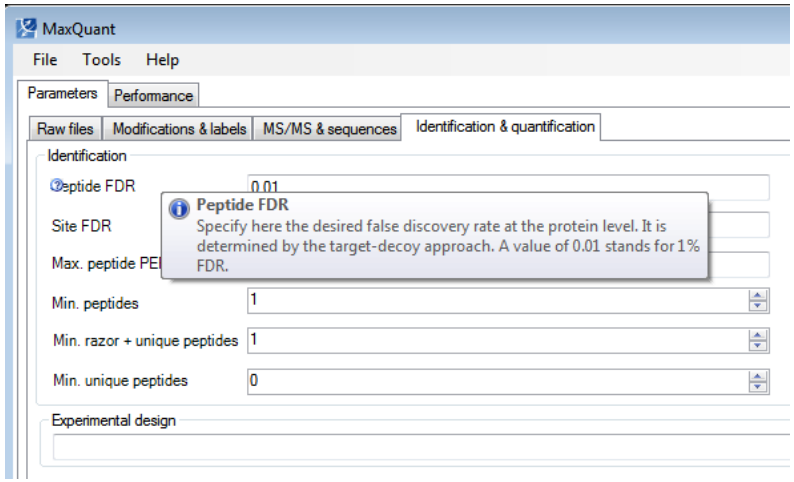


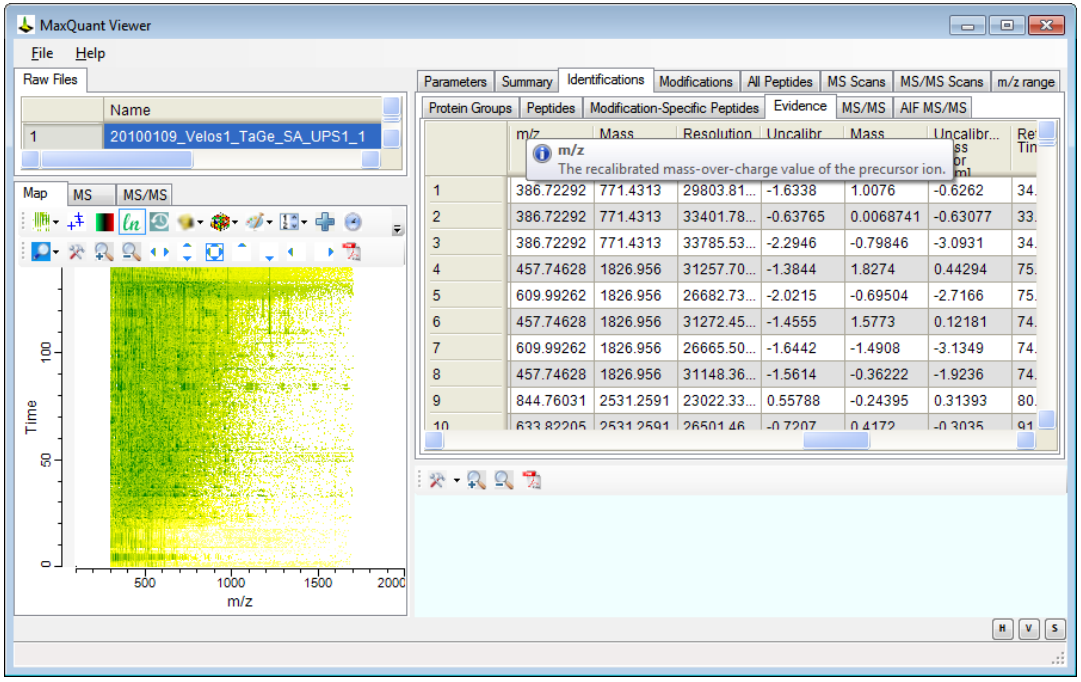
www.maxquant.org

groups.google.com/group/maxquant-list



Usability, documentation, software quality





Getting started with MaxQuant

Welcome to the MaxQuant computational proteomics workflow. We hope that you will enjoy your time you will be spending with processing your data and that MaxQuant delivers some helpful results. Currently we support the following instrument types:

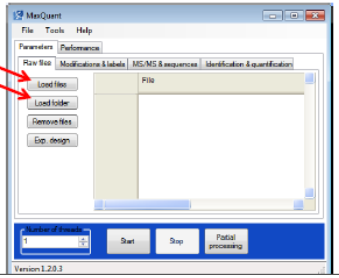
1. Thermo Velos
2. Thermo Orbitrap
3. Thermo FT
4. Thermo Exactive

Many labeling technologies are supported. In fact, MS-level quantification labels can be freely configured. Supported sample labeling quantification tectniques include:

1. SILAC
2. Label-free quantification
3. Dimethyl
4. 18O
5. ICAT
6. ICPL
7. mTRAQ

In case you are a first time user you might be worried by the many options and parameters that one can set in the user interface. In that case we have good news for you. In almost all use cases the standard values of most parameters are fine and you only need to adjust a small number of factors. Typically there is only little information that you need to provide. Every parameter in the interface has context help which you obtain by moving the mouse pointer to the beginning of the text string for this parameter and clicking on the question mark that will appear. Here are seven steps that you typically have to go through before MaxQuant can process your data:

1. Specify the raw files that you want to process with MaxQuant.



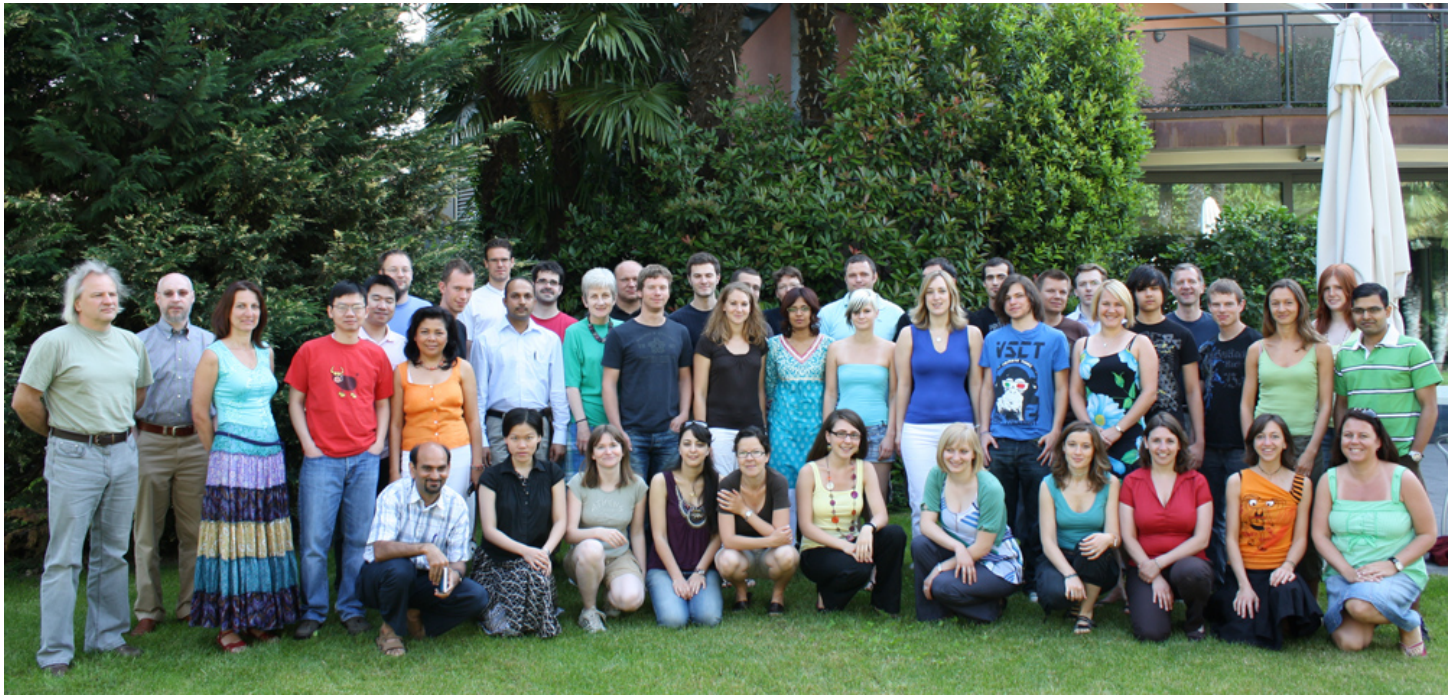
Acknowledgements

Matthias Mann

Nadin Neuhauser
Richard Scheltema
Christoph Schaab

Christian Luber

All Mann lab members



Thank you for your attention

http://www.maxquant.org

http://groups.google.com/group/maxquant-list
http://groups.google.com/group/andromeda-list
http://groups.google.com/group/perseus-list

cox@biochem.mpg.de