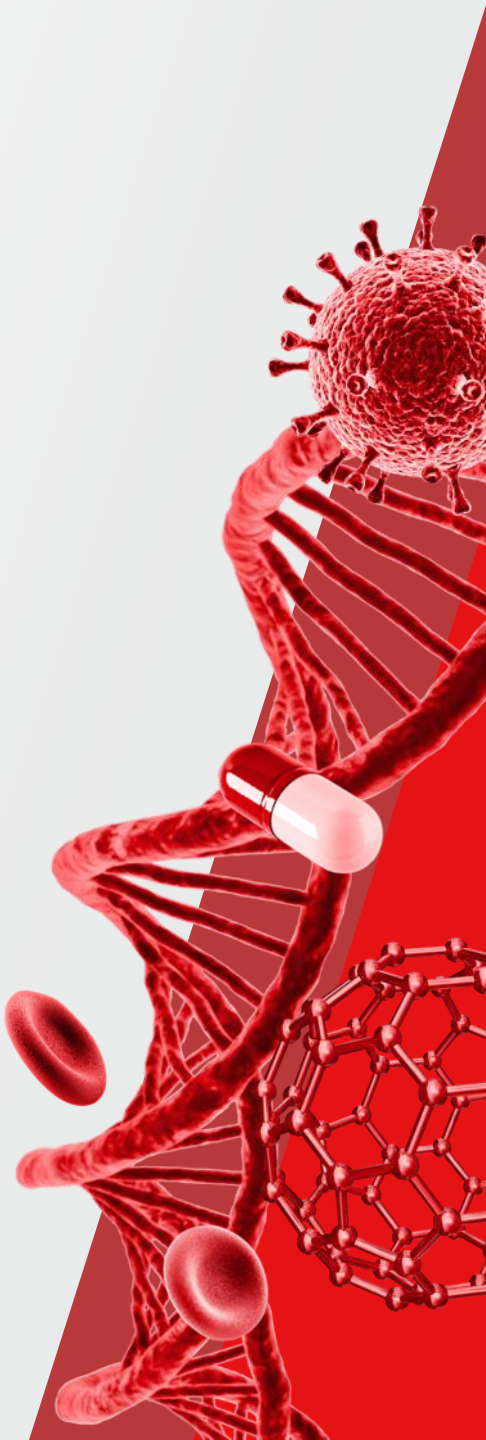
Proteome Discoverer Workshop

Hradec Králové

Ing. Michaela Scigelova, Ph.D.
LSMS Factory Support Groups
Bremen, Germany



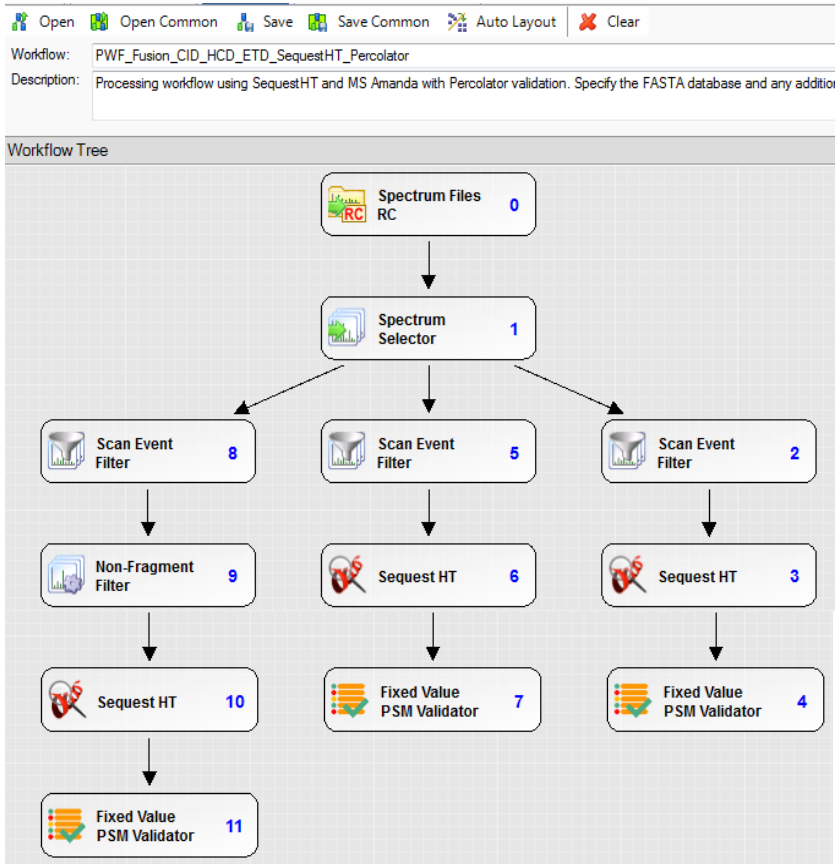
The world leader in serving science

Agenda

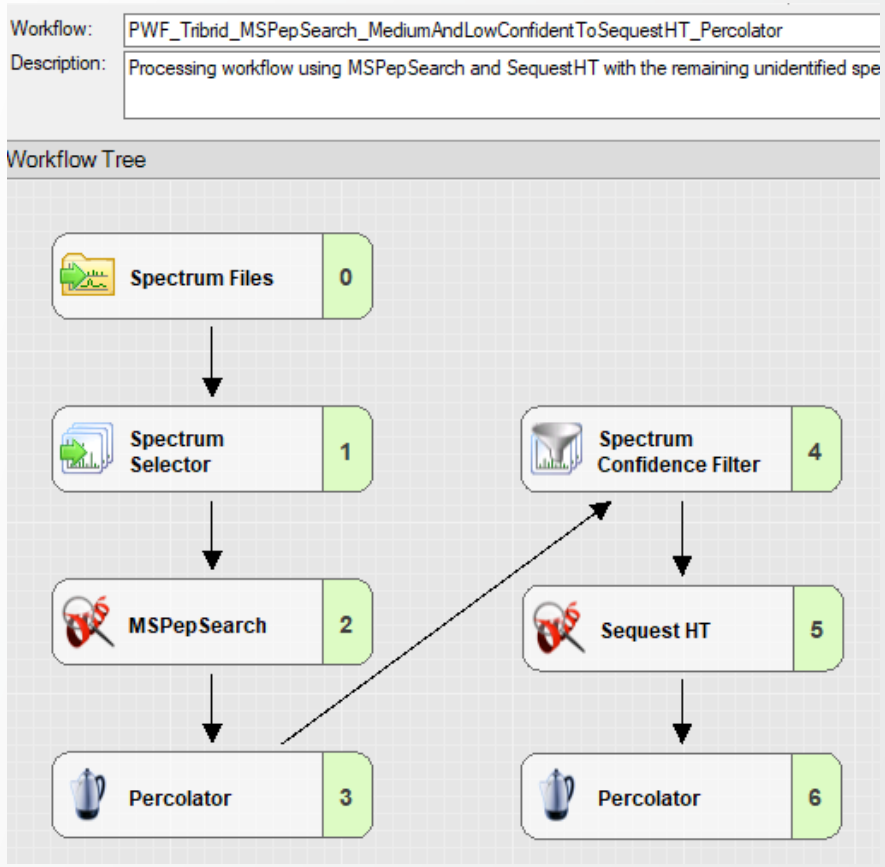
- What's new
  - Enhanced workflow capabilities
  - Spectral libraries
  - INFERYS
  - CHIMERYS
- TMT quantification
  - TMTpro 18plex
  - PSM filters for quan
  - Normalisation and scaling
  - Multiplexed TMT study design

Enhanced Workflow Capabilities

- Branched

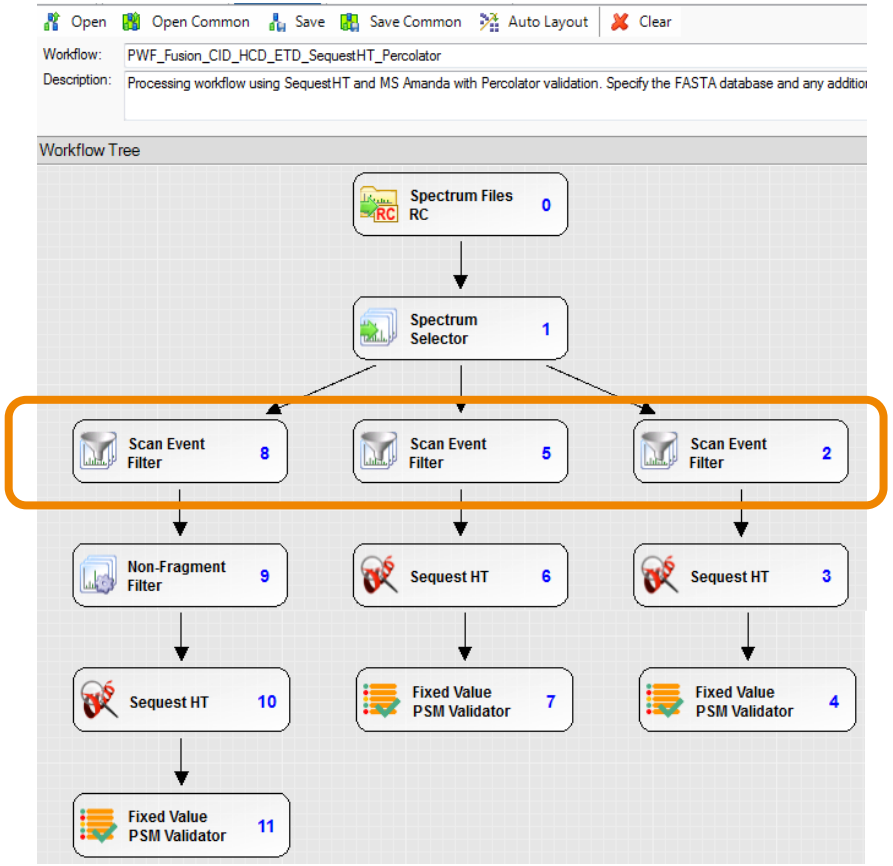


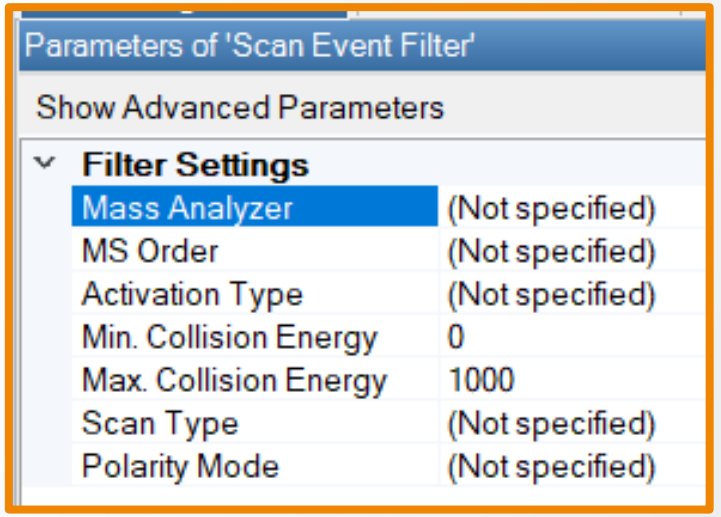
Iterative



Branched WF

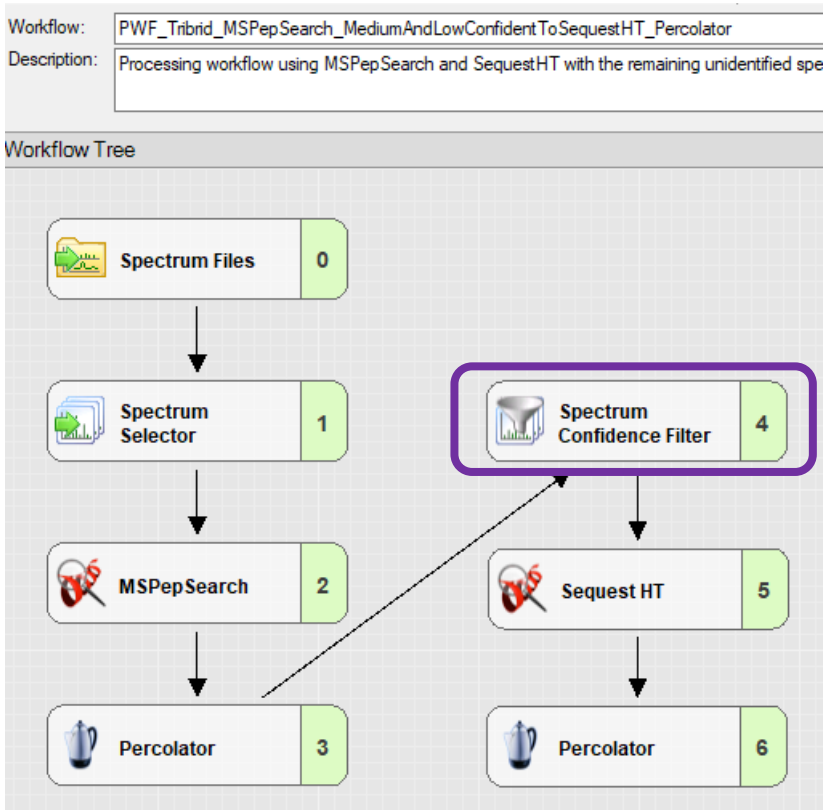
- Use Scan Event Filter node

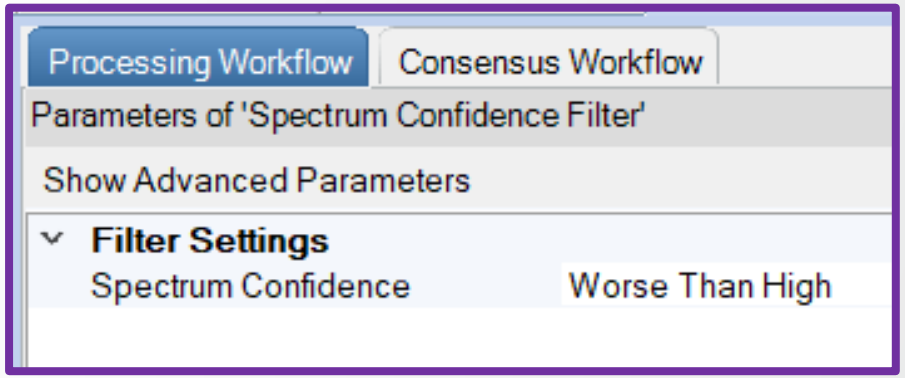




Iterative WF

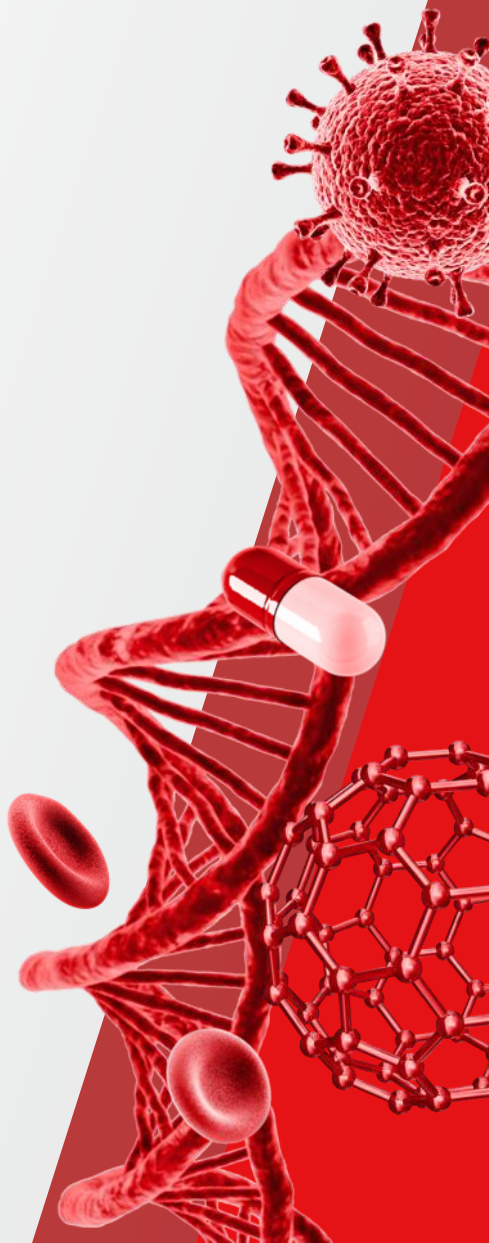
- Use Spectrum Confidence Filter node
  - Sends selected spectra for the next round of processing





ThermoFisher SCIENTIFIC

Spectral Libraries



The world leader in serving science

Spectral Libraries in PD

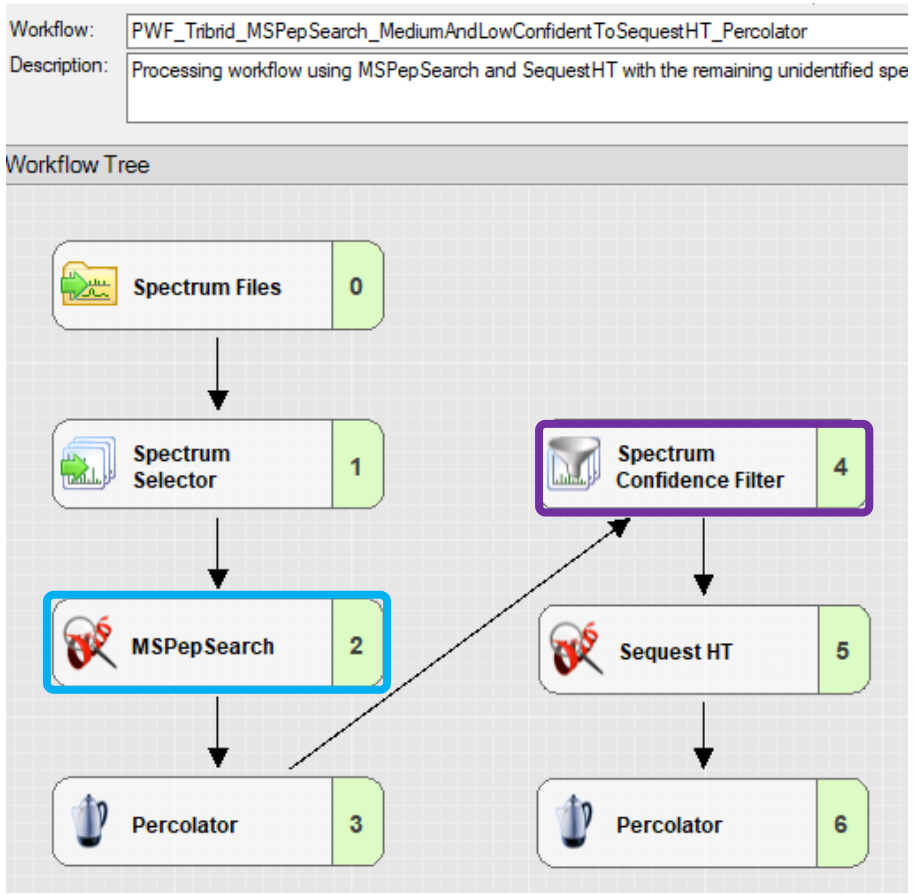
- Library search engine
  - Uses fragment abundances in the correlation of the experimental and library spectra

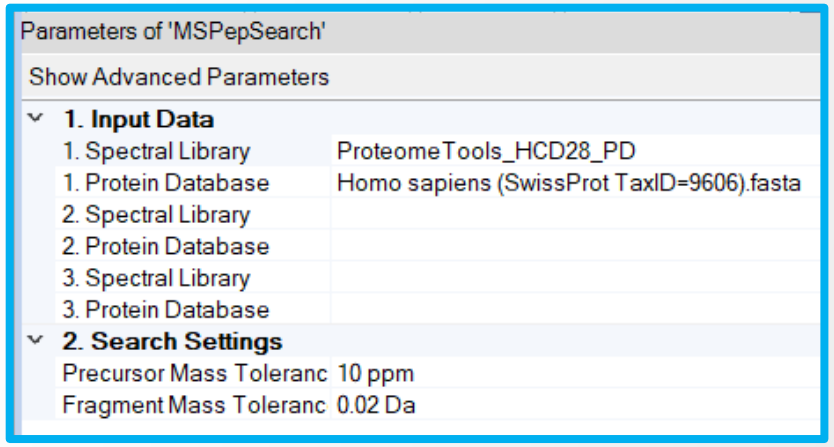
- Complements database search
  - Fast, multithreaded in PD 2.5, more sensitive
  - With confidence estimate based on decoy library search (FDR)

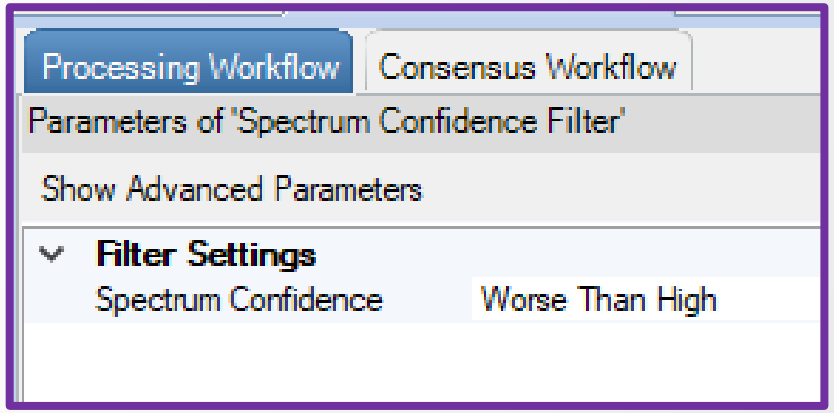
- Libraries can be directly downloaded in PD Administration
  - **NIST** – large number of spectra, various modifications
  - **ProteomeTools** – libraries of synthetic peptides, some modified peptides included, acquired at various CE settings

Spectral Library Search

- MSPepSearch deployed as part of an iterative WF

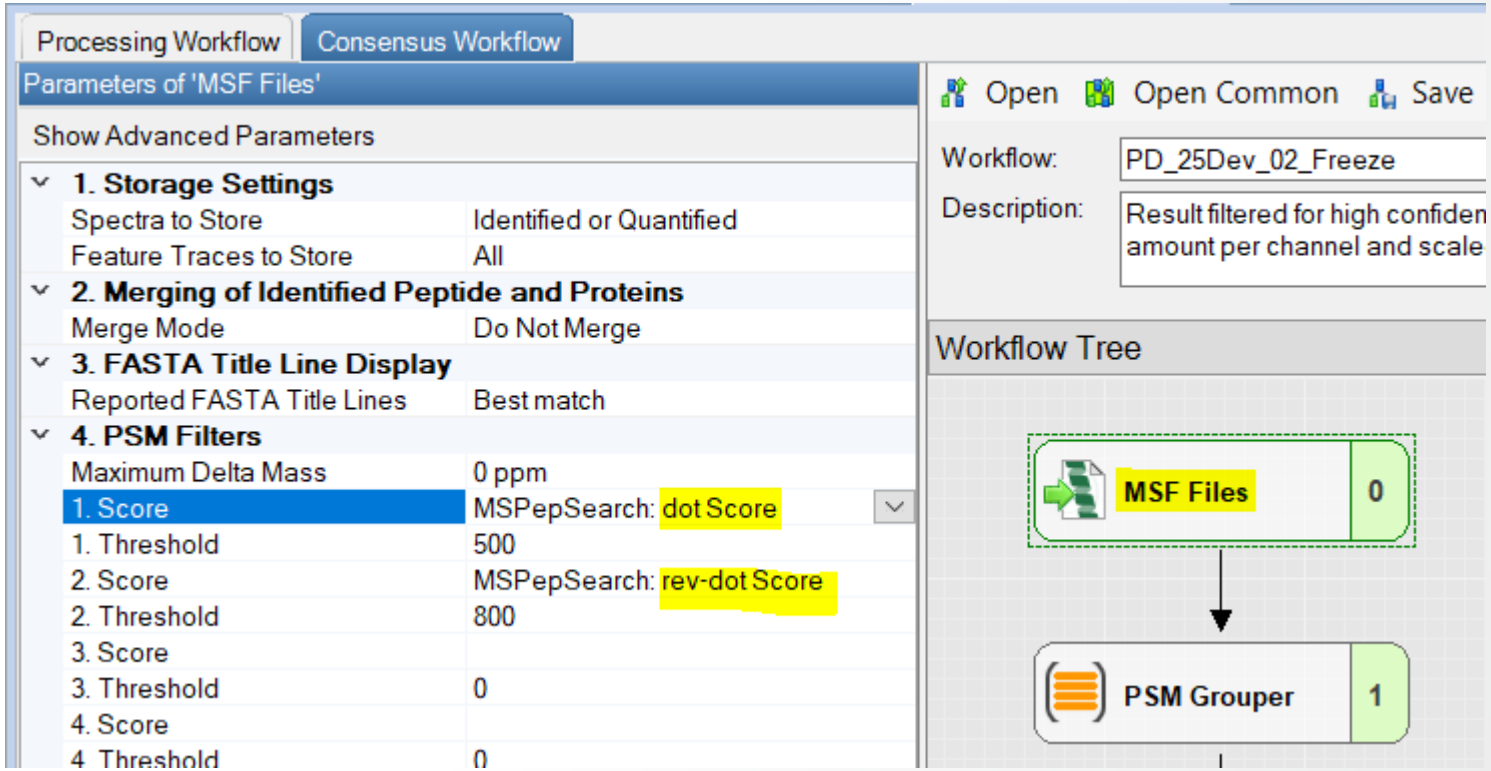




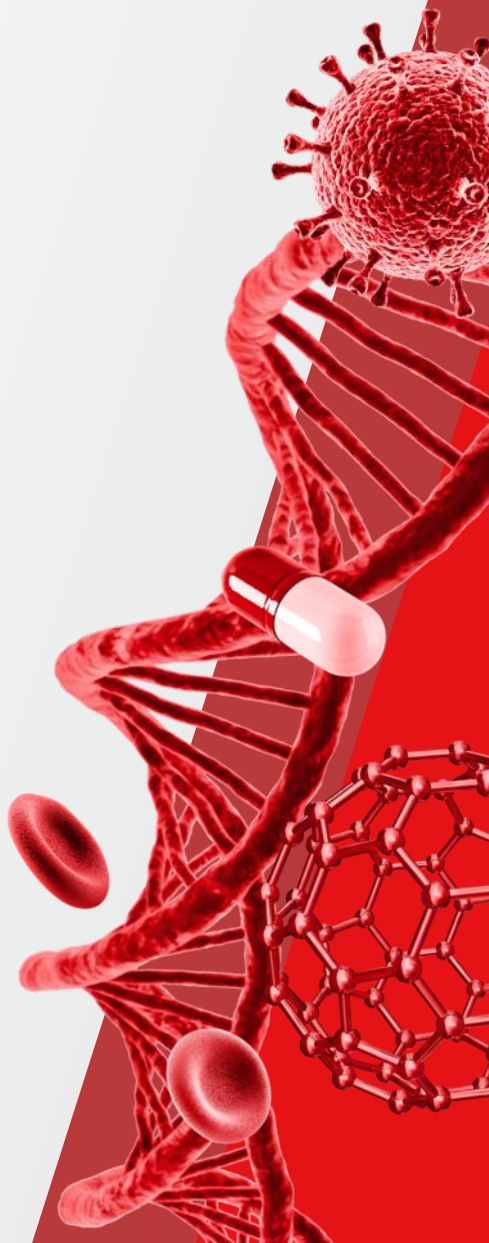


Interpreting Library Search Results

- Empirically derived score thresholds (my own; not a general recommendation)
  - dot Score >= 500
  - rev-dot Score >= 800
- Set as PSM filters in MSF Files node



INFERYS - Predicted Spectral Libraries



The world leader in serving science

Predicted spectral libraries

- Deep learning-based model for peptide MS/MS spectral prediction
- Predicts peptide spectrum (including relative intensities of fragments)
- Remarkable accuracy

Generating high-quality libraries for DIA-MS with empirically-corrected peptide predictions

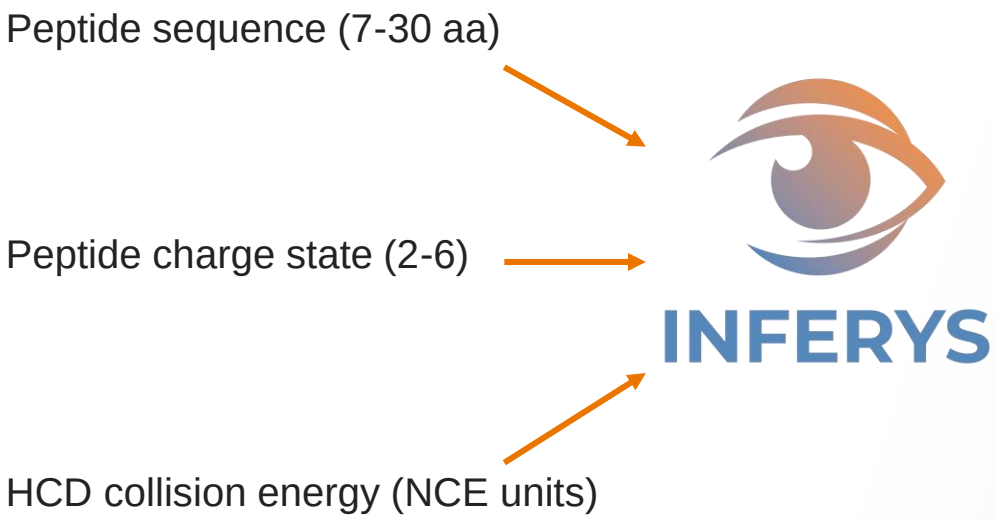
Brian C. Searle, Kristian E. Swearingen, Christopher A. Barnes, Tobias Schmidt, Siegfried Gessulat, Bernhard Kuster, Mathias Wilhelm

doi: https://doi.org/10.1101/682245

Now published in *Nature Communications* doi: 10.1038/s41467-020-15346-1

Predicted Spectral Libraries

- Deep learning-based model for peptide MS/MS spectra prediction



Predicted Spectral Libraries

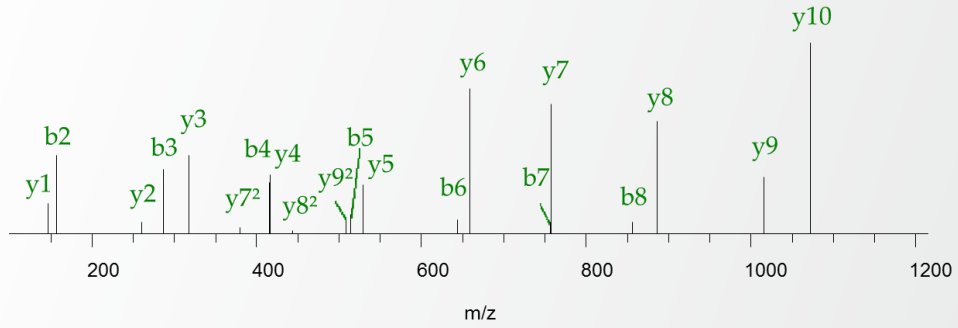
- Deep learning-based model for peptide MS/MS spectra prediction

VGEEVEIVGIK

z = 2+

NCE = 26

INFERYS

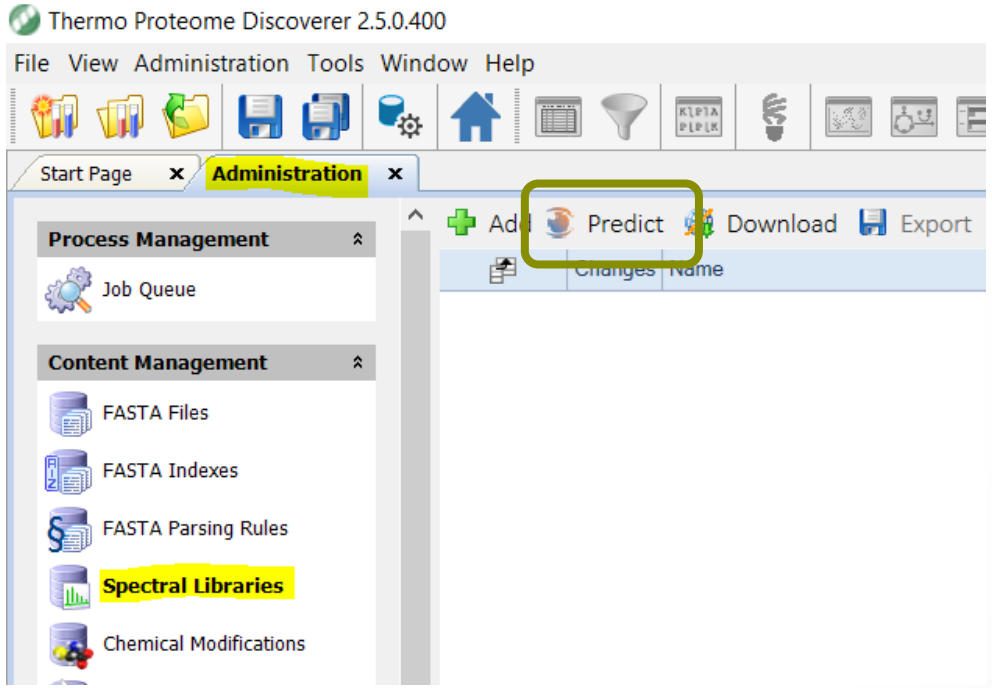


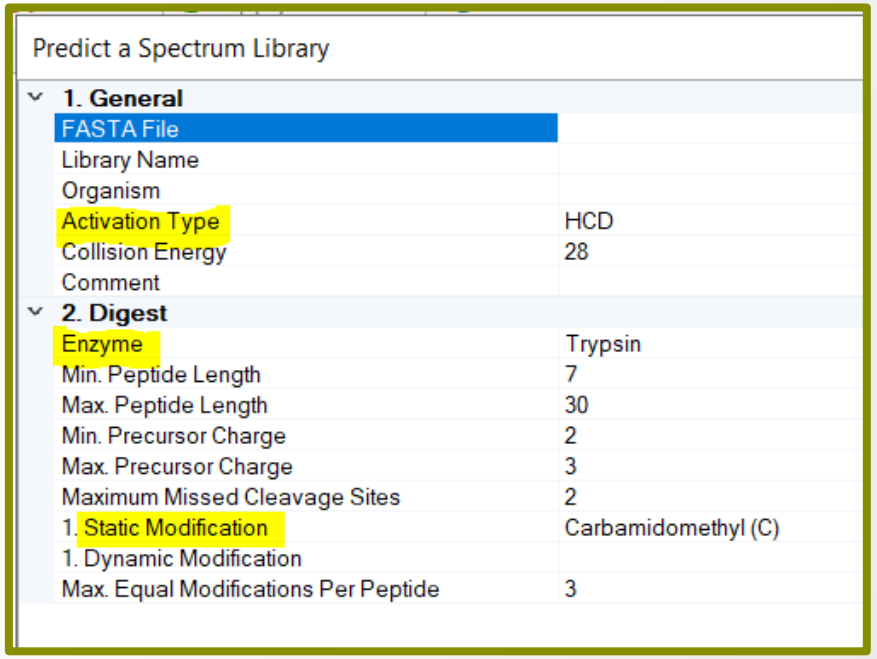
Predicted MS/MS spectrum

Predicted Spectral Libraries

- Administration → Maintain Spectrum Libraries → Predict

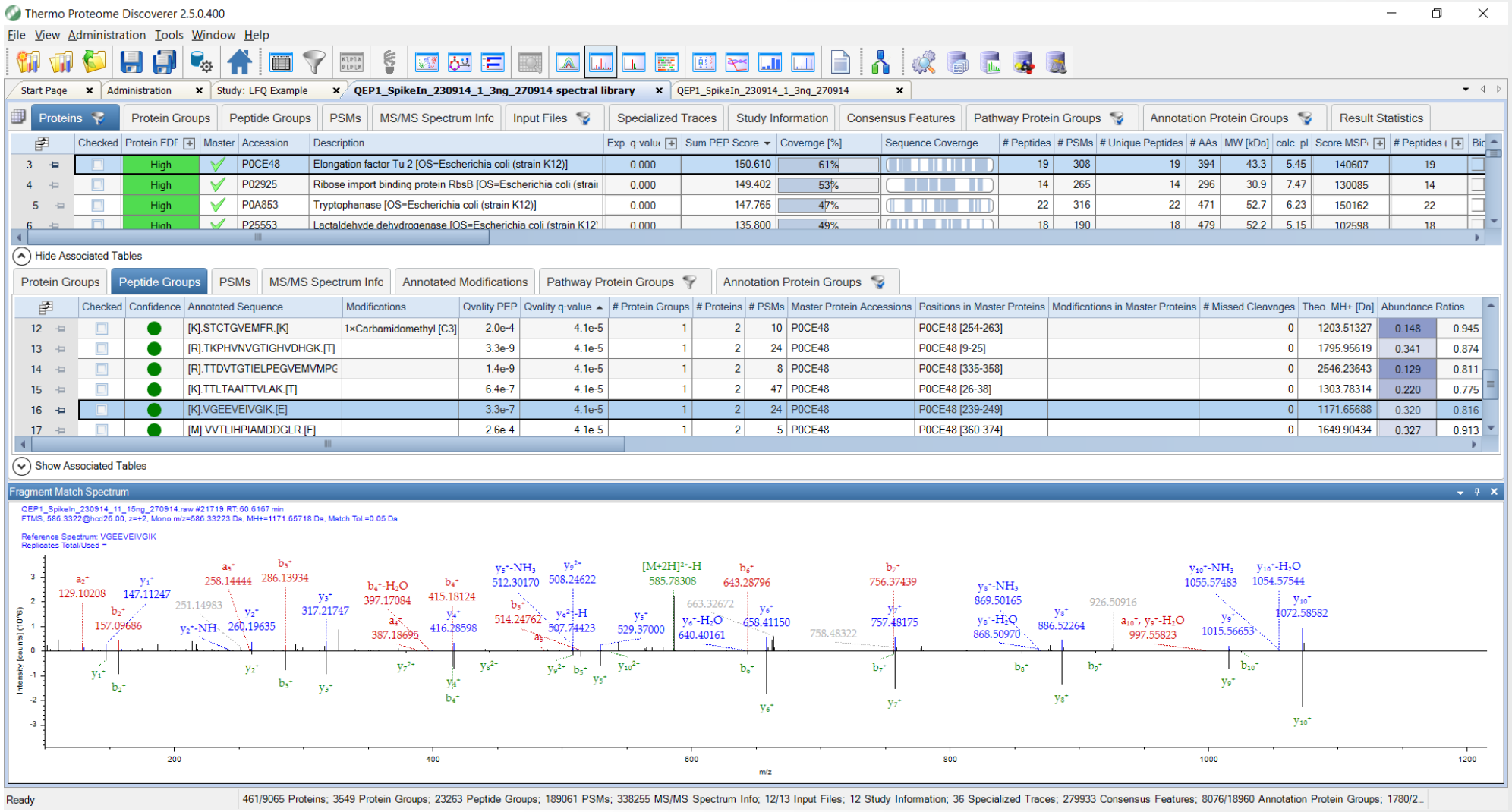
- Create a predicted spectral library for all peptides in a FASTA file
- Use predicted libraries for searches with MSPepSearch node





Search Result with Predicted Library

- Match between experimental and predicted spectrum



Caveats

- Prediction of spectral libraries has a **limited practical use**
  - Recommended for small FASTA files
  - O.K. to characterize a recombinant protein or a protein complex
  - High consumption of disc space

Caveats

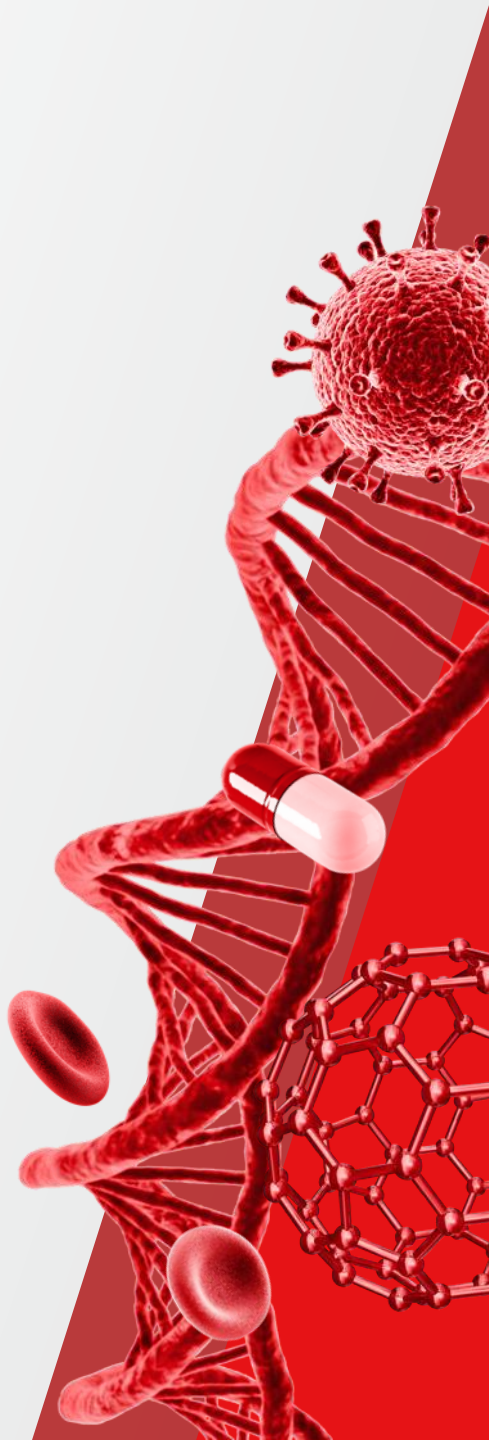
- High consumption of disk space

- *Example: working with human SwissProt-derived library*
  - Predicted library size 11 - 12 GB
  - Decoy DB of approx. the same size needed
  - When a library is being predicted, there are intermediate files (the prediction queries and the prediction result files) being saved on the system (100 GB free disc space needed)
  - The translation of the prediction format to the NIST binary format required for the searches also needs a lot of space
  - PD stores the annotated reference spectra in a separate library file for a quick display. This file is 10-20 GB

- It takes about 1 day

ThermoFisher SCIENTIFIC

INFERYS - Rescoring

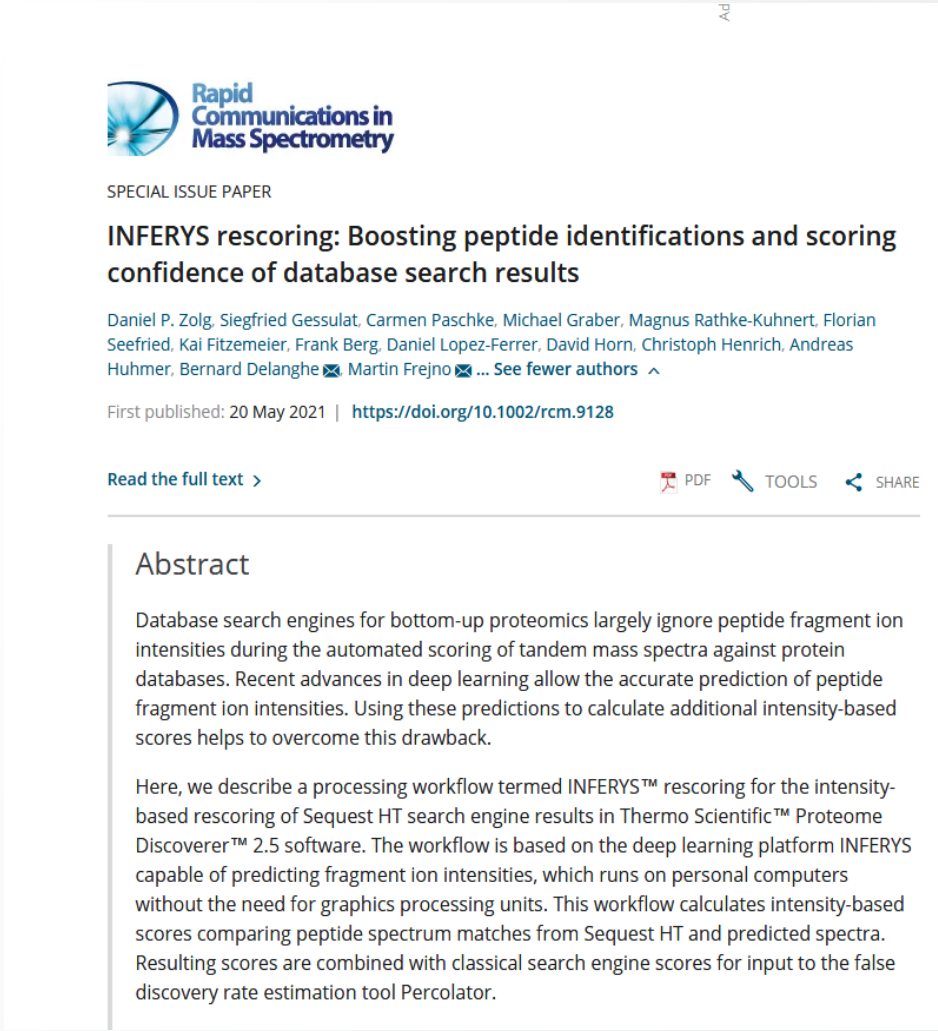


The world leader in serving science

Prediction-based rescoring for improved confidence

- **INFERYS Rescoring** node
  - predicts MS/MS spectra for PSMs identified by Sequest HT
  - subsequently compares the predicted and experimental spectra
- Updated **Percolator** node
  - includes new features for improved peptide ID confidence

https://doi.org/10.1002/rcm.9128



Rapid Communications in Mass Spectrometry

SPECIAL ISSUE PAPER

INFERYS rescoring: Boosting peptide identifications and scoring confidence of database search results

Daniel P. Zolg, Siegfried Gessulat, Carmen Paschke, Michael Graber, Magnus Rathke-Kuhnert, Florian Seefried, Kai Fitzemeier, Frank Berg, Daniel Lopez-Ferrer, David Horn, Christoph Henrich, Andreas Huhmer, Bernard Delanghe, Martin Frejno ... See fewer authors

First published: 20 May 2021 | https://doi.org/10.1002/rcm.9128

Read the full text › PDF TOOLS SHARE

Abstract

Database search engines for bottom-up proteomics largely ignore peptide fragment ion intensities during the automated scoring of tandem mass spectra against protein databases. Recent advances in deep learning allow the accurate prediction of peptide fragment ion intensities. Using these predictions to calculate additional intensity-based scores helps to overcome this drawback.

Here, we describe a processing workflow termed INFERYS™ rescoring for the intensity-based rescoring of Sequest HT search engine results in Thermo Scientific™ Proteome Discoverer™ 2.5 software. The workflow is based on the deep learning platform INFERYS capable of predicting fragment ion intensities, which runs on personal computers without the need for graphics processing units. This workflow calculates intensity-based scores comparing peptide spectrum matches from Sequest HT and predicted spectra. Resulting scores are combined with classical search engine scores for input to the false discovery rate estimation tool Percolator.

INFERYS Rescoring Node

Additional features calculated as input into Percolator

- Search against FASTA file
- Produce a list of PSMs from the target and decoy database searches
- Produce up to rank 10 hits per spectrum

- Determine optimal collision energy for input into INFERYS algorithm
- Calculate all predicted spectra for all target and decoy hits
- Calculate spectral angle and other figures-of-merit for correlation of predicted and measured spectra

- Use the standard 35 features
- If INFERYS Rescoring node is used, add 15 more features
- Calculate FDR thresholds using SVM score calculated from all input features
- For information about Percolator, read Matthew The *et* al, JASMS 2016

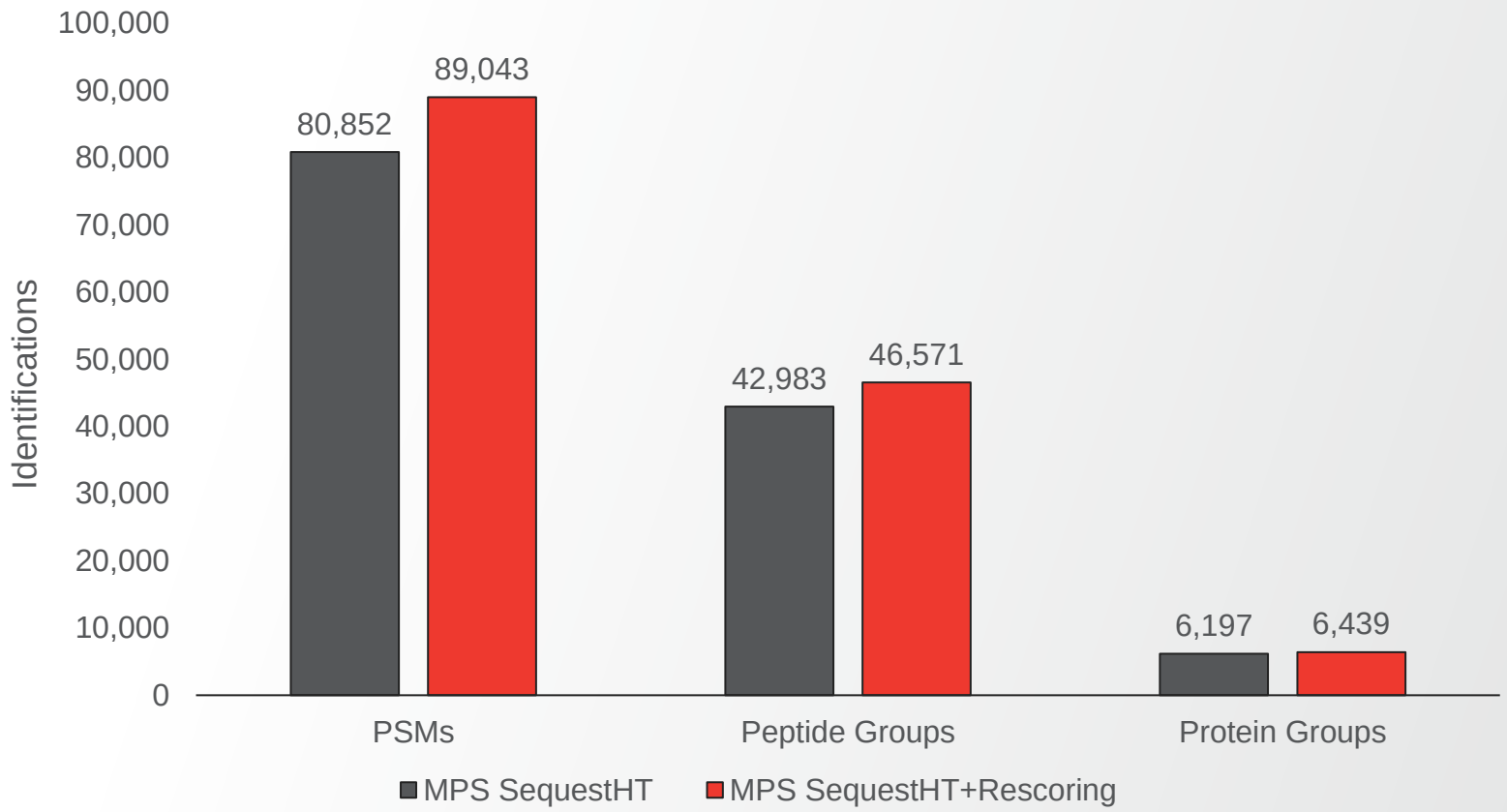
Results – Standard bottom-up proteomics data set

200 ng HeLa Thermo Scientific™ Obitrap Exploris™ 480 MS with FAIMS CV50 CV70

- Trypsin digest

- 10% increase in PSMs
- 8% more unique peptides
- 4% more proteins

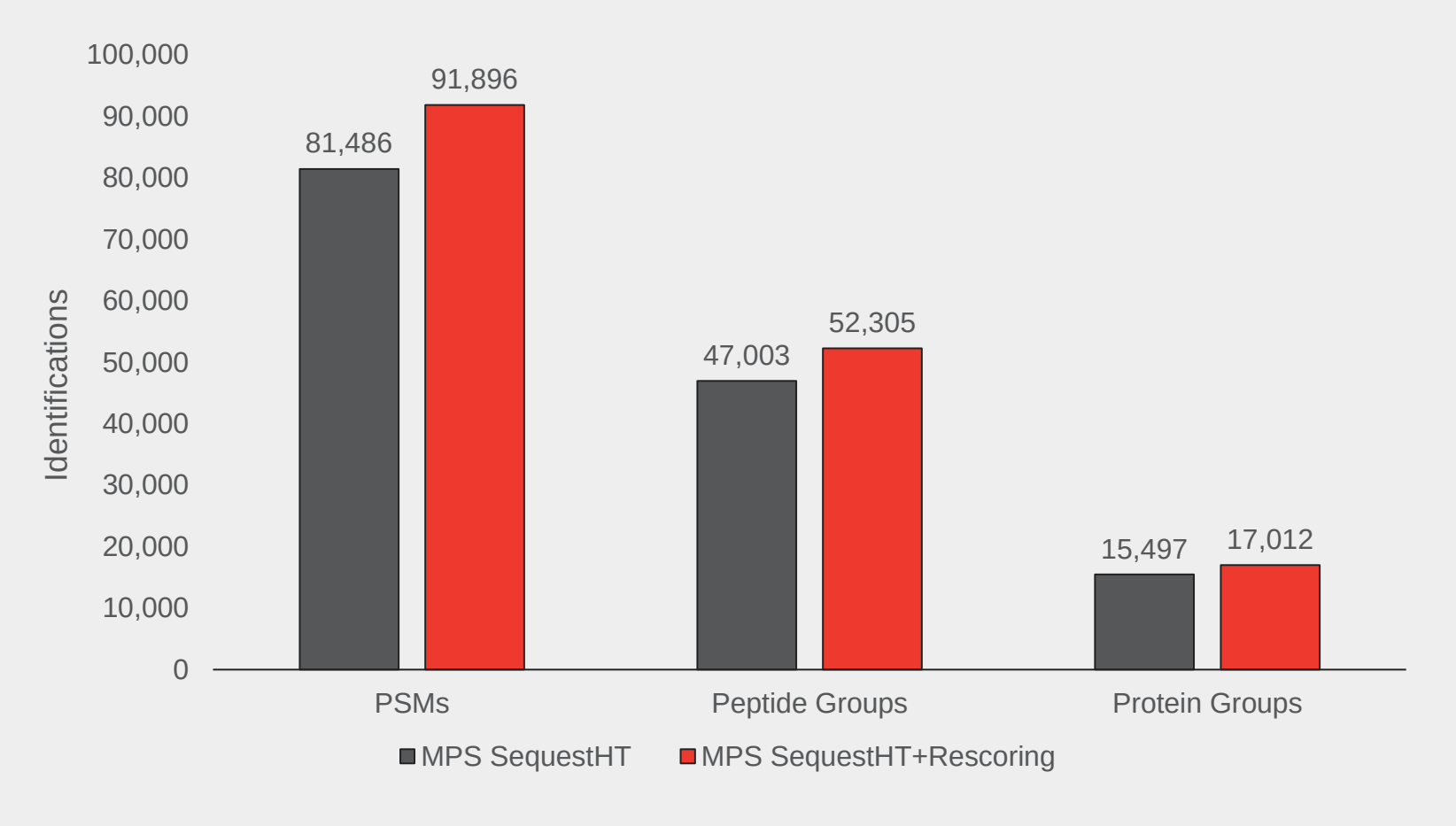
- Rescues short peptides and peptides where classical scores fail to separate targets from decoys



Results – Standard bottom-up proteomics data set

Metaproteomics – Human stool samples (*Rechenberger et al., Proteomes. 2019)*

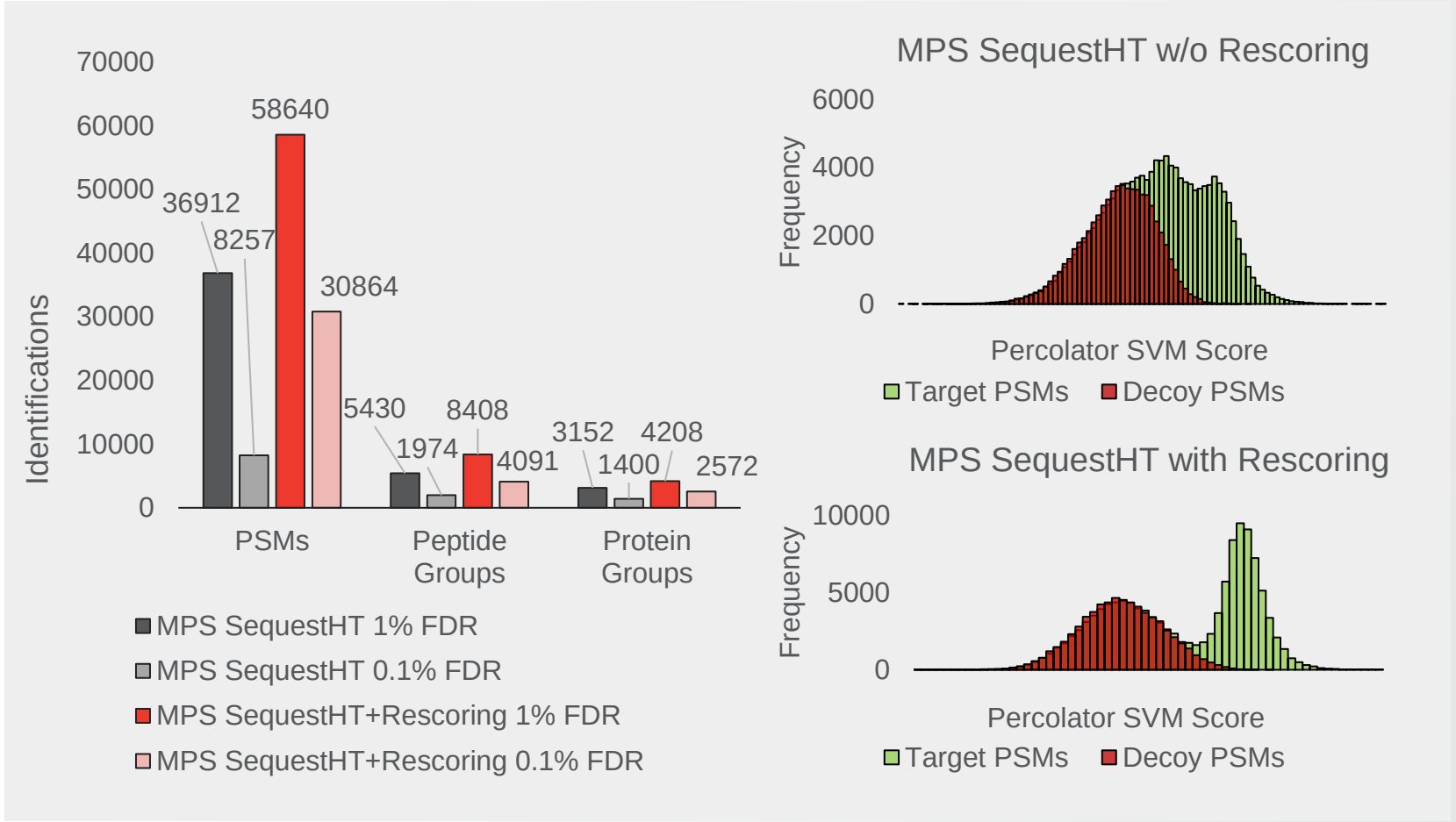
- Trypsin digest
- Complex sample with huge search space
- 13% increase in PSMs
- 11% more peptides
- 10% more proteins
- Rescues short peptides and peptides where classical scores fail to separate targets from decoys



Results – Immunopeptidomics data set

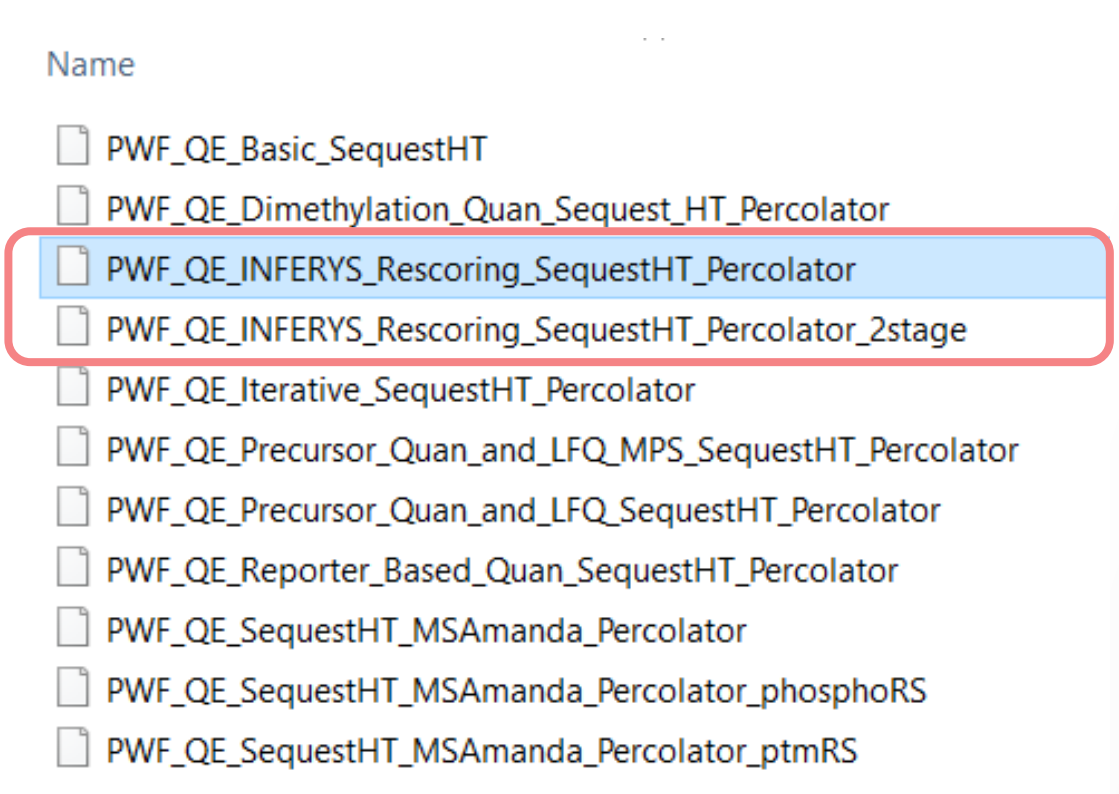
HLA Class I data set – Patient derived melanoma cell line (*Chong. et al, Nat. Com. 2020*)

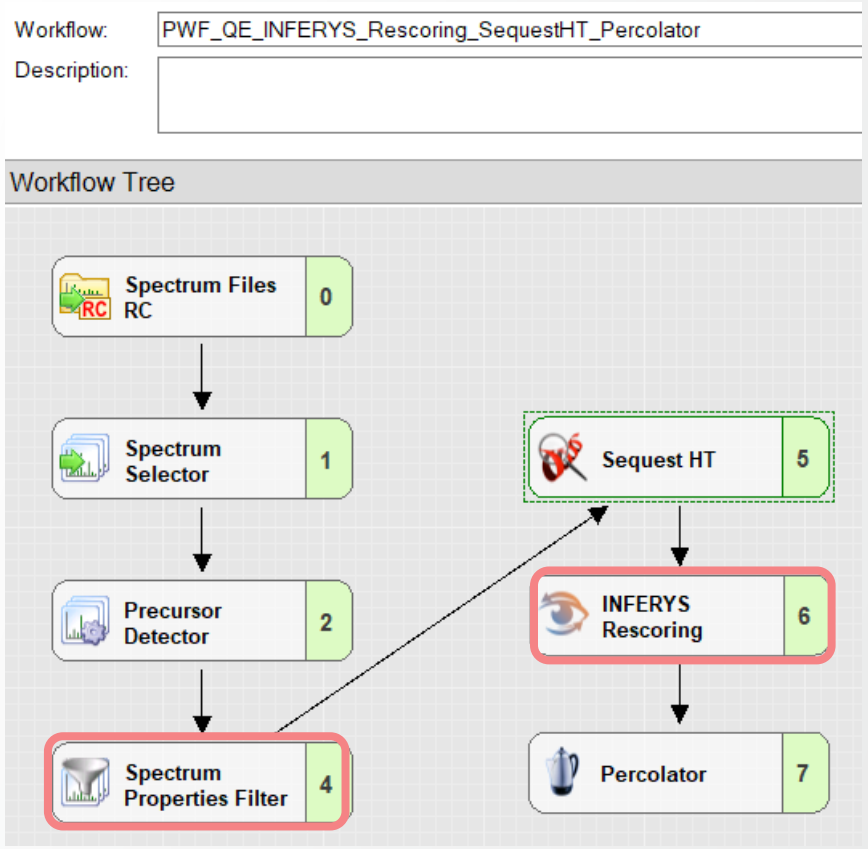
- No enzyme search
- Huge search space and very similar peptide properties
- 59% increase in PSMs
- 55% more peptides
- 34% more proteins
- Reduces peptide loss at 0.1% compared to 1% FDR from 64% to 25%



Default Processing Workflows

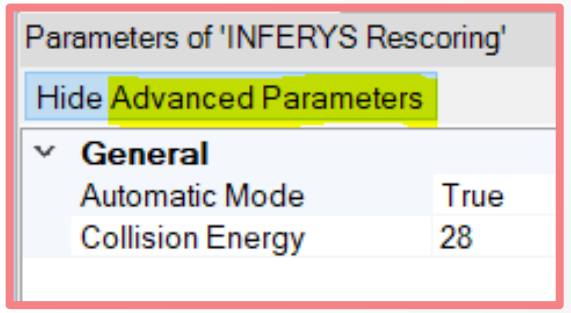
- Q Exactive

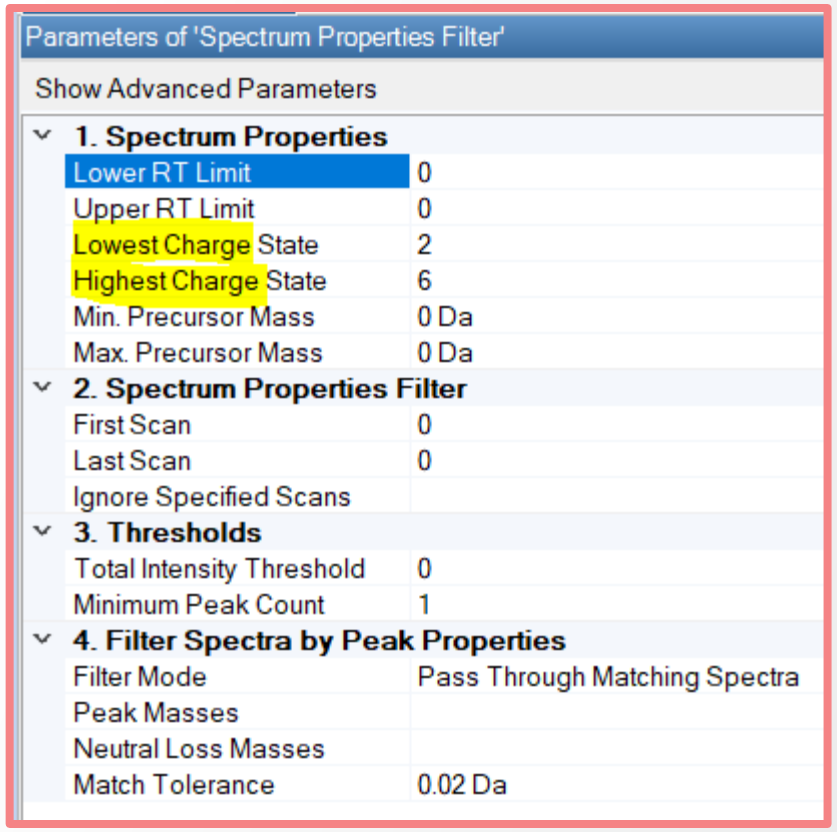




Default Processing Workflows

- INFERYS node
  - Define collision energy in “advanced parameters”
- Must include Spectrum Properties Filter node





Summary – INFERYS Rescoring

- **Significant increase in IDs**

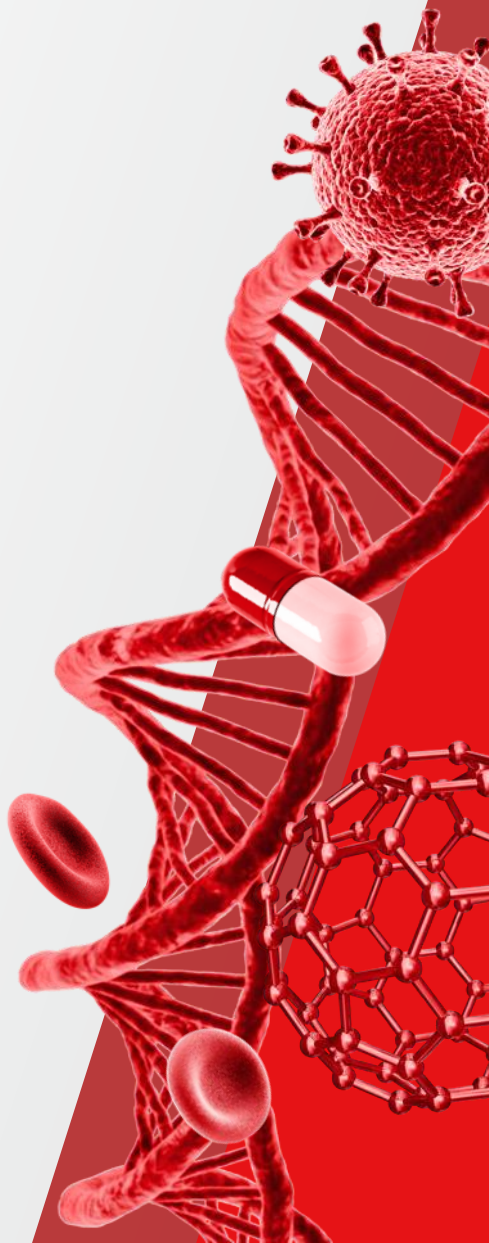
- Useful for extremely large search space in particular
  - no-enzyme searches
  - metaproteomics

Caveats

- Input from .raw file
- High res HCD spectra
- Charge state 2-6
- Only with Sequest HT (does not work with Mascot)
- Not trained for TMT-labeled peptides

ThermoFisher SCIENTIFIC

CHIMERYS



The world leader in serving science

CHIMERYS

Deconvolutes chimeric spectra based on predicted fragment ion intensities.

A cloud-native search algorithm that uses accurate predictions of peptide fragment ion intensities and retention times provided by the deep learning framework INFERYS 2.0.

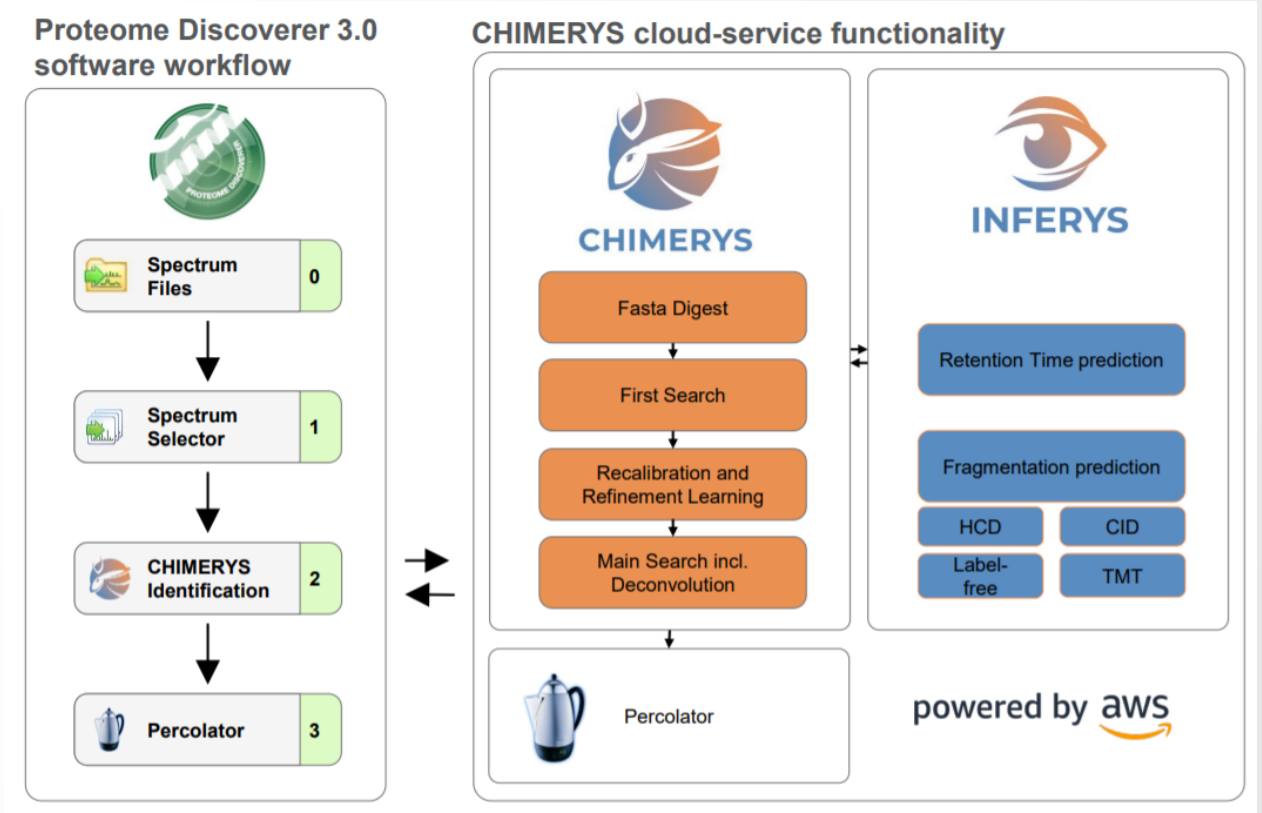
- Artificial intelligence-powered search algorithm
- Implemented in PD 3.0 (release February 2022)
- 2x peptide IDs compared to Sequest HT alone
- Protein coverage increase by 2.5-fold
- More than 80% identification rate

MSAID thermo scientific

CHIMERYS

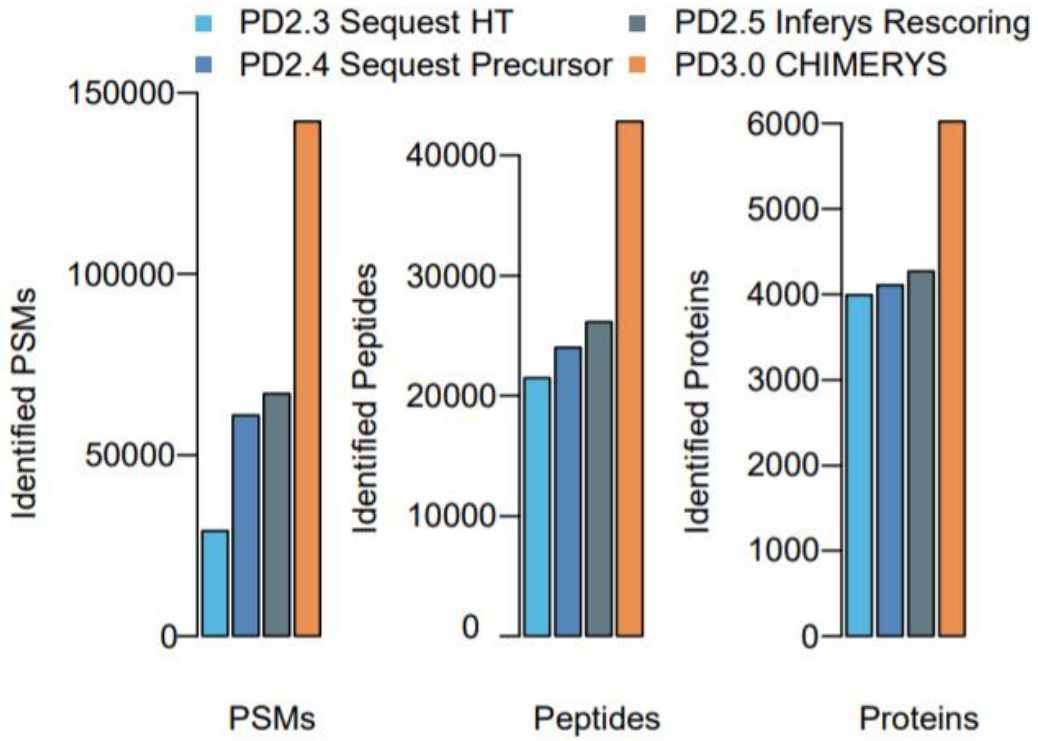
Deconvolutes chimeric spectra based on predicted fragment ion intensities

- All candidates in the isolation window of a given tandem mass spectrum are considered simultaneously
- Compete for measured fragment ion intensity in one concerted step
- CHIMERYS aims to explain as much measured intensity with as few candidate peptides as possible, resulting in the **deconvolution of chimeric spectra**
- Peptide spectrum match (PSM)-level false discovery rate (FDR)-control is performed using Percolator

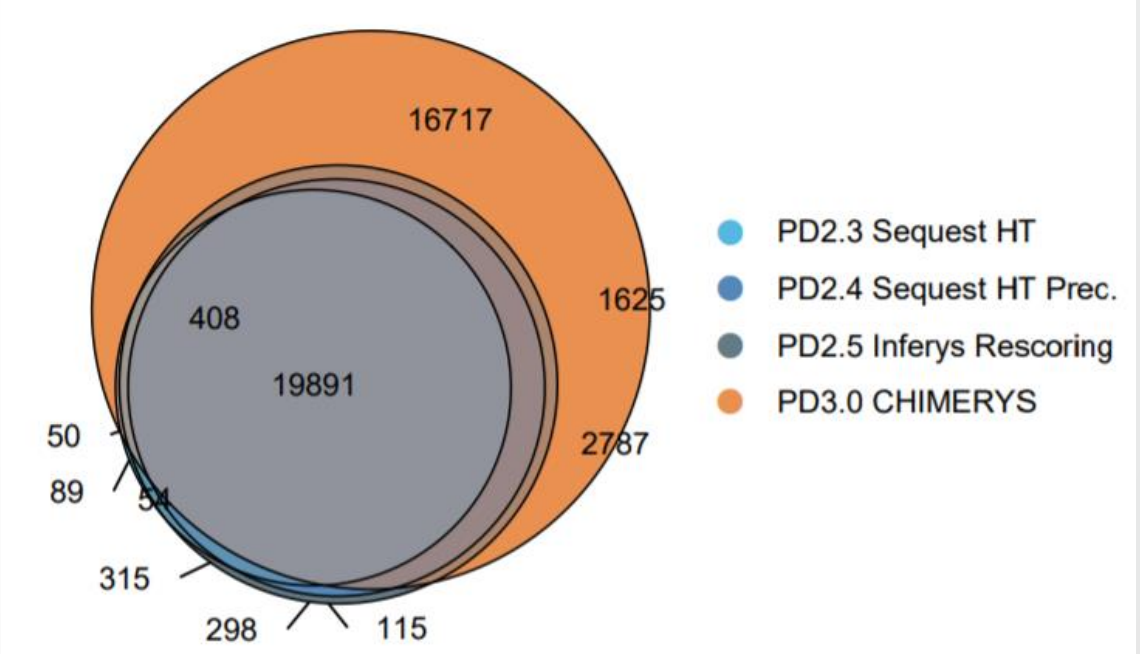


Peptide and Protein IDs

HeLa cell lysate, 1-hour gradient, Thermo Scientific™ Orbitrap Exploris™ 480



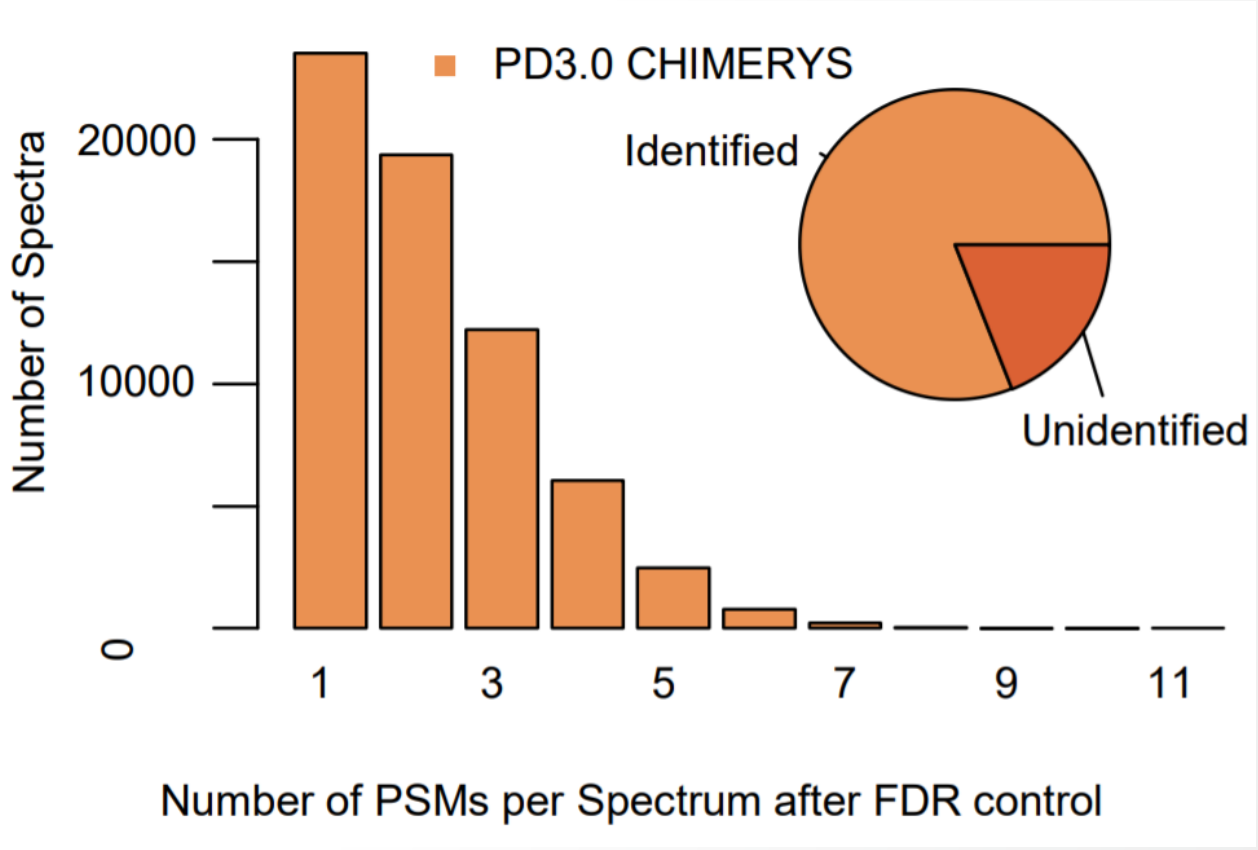
Overlap of peptide identifications



Extent of Chimeric Spectra Problem

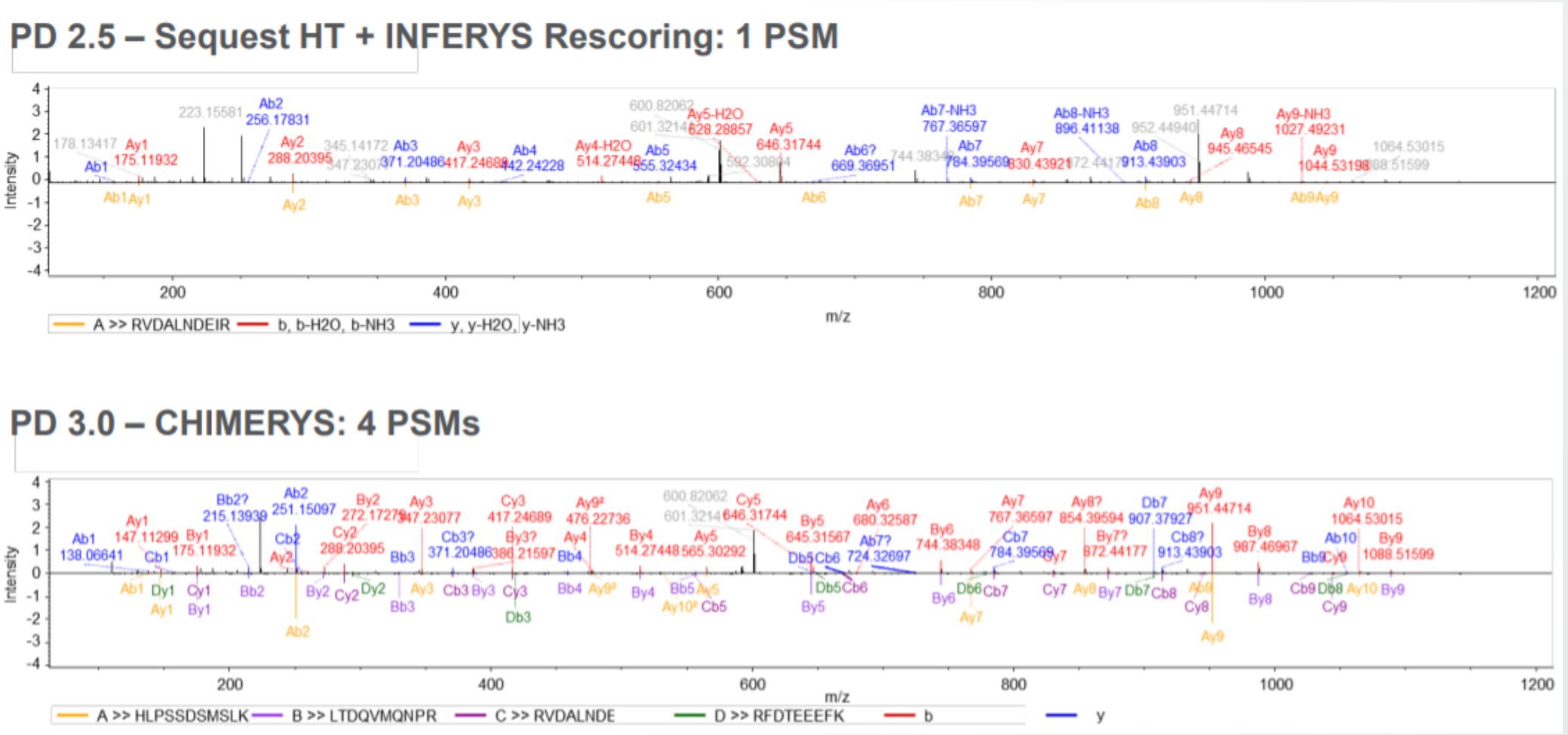
Number of PSMs per spectrum and identification rate achieved by CHIMERYS

- More than 80% identification rate



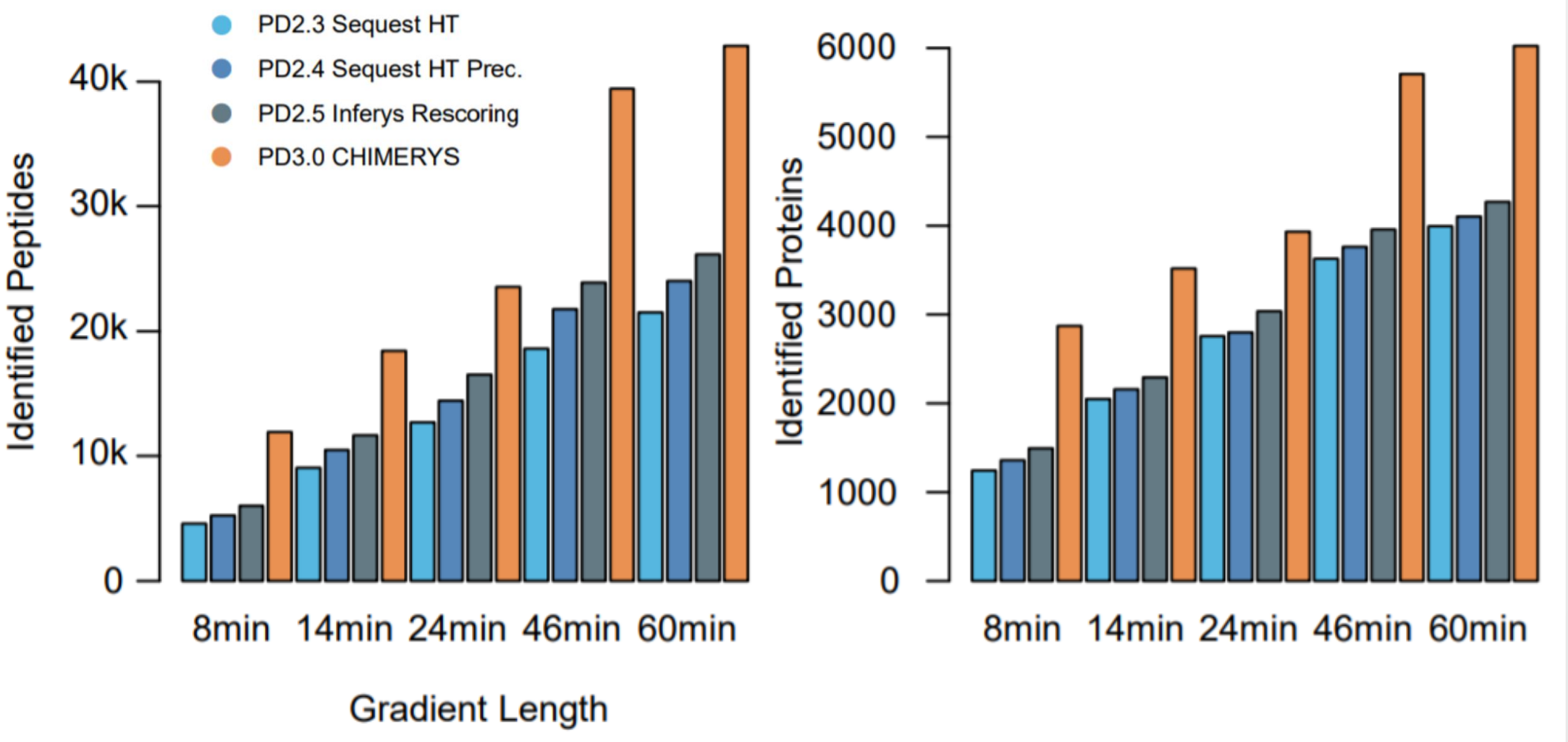
Proteome Discoverer Spectrum Viewer

Visualise the proportional contributions of the individual peptides in a mirror plot



Increased Throughput

Same number of peptides and protein groups in 1/3 of the measurement time

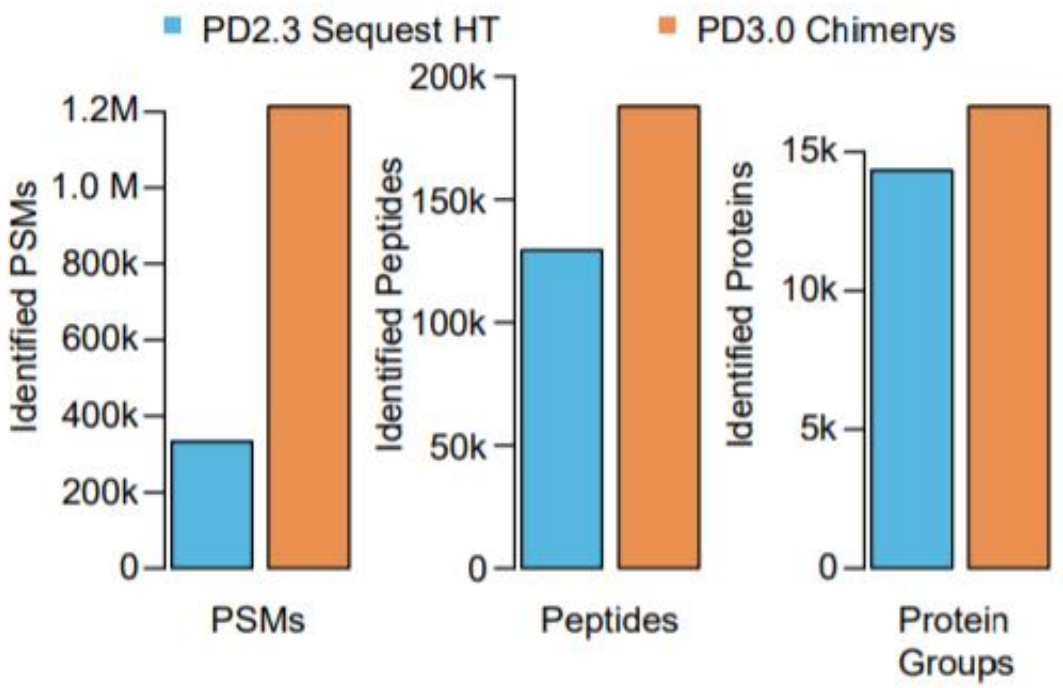


Applicability

Organisms from all kingdoms of life as well as less complex samples like body fluids

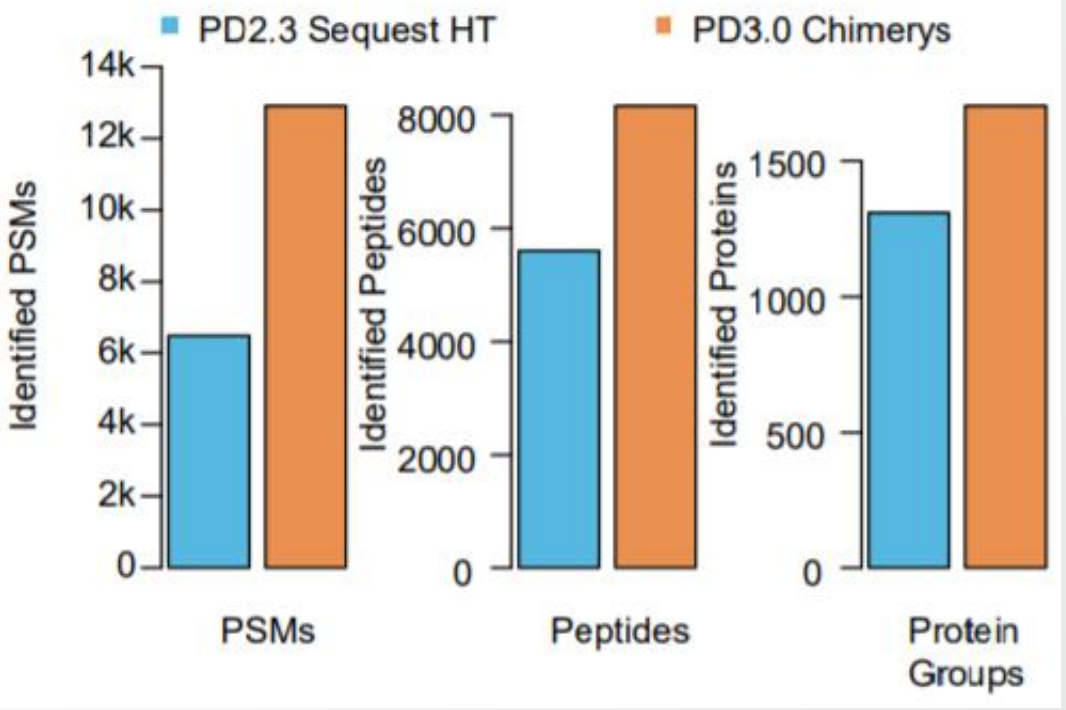
Fractionated *Arabidopsis thaliana* proteome

PRIDE Project PXD019483



A single 30 min urine proteome file

PRIDE Project PXD015087



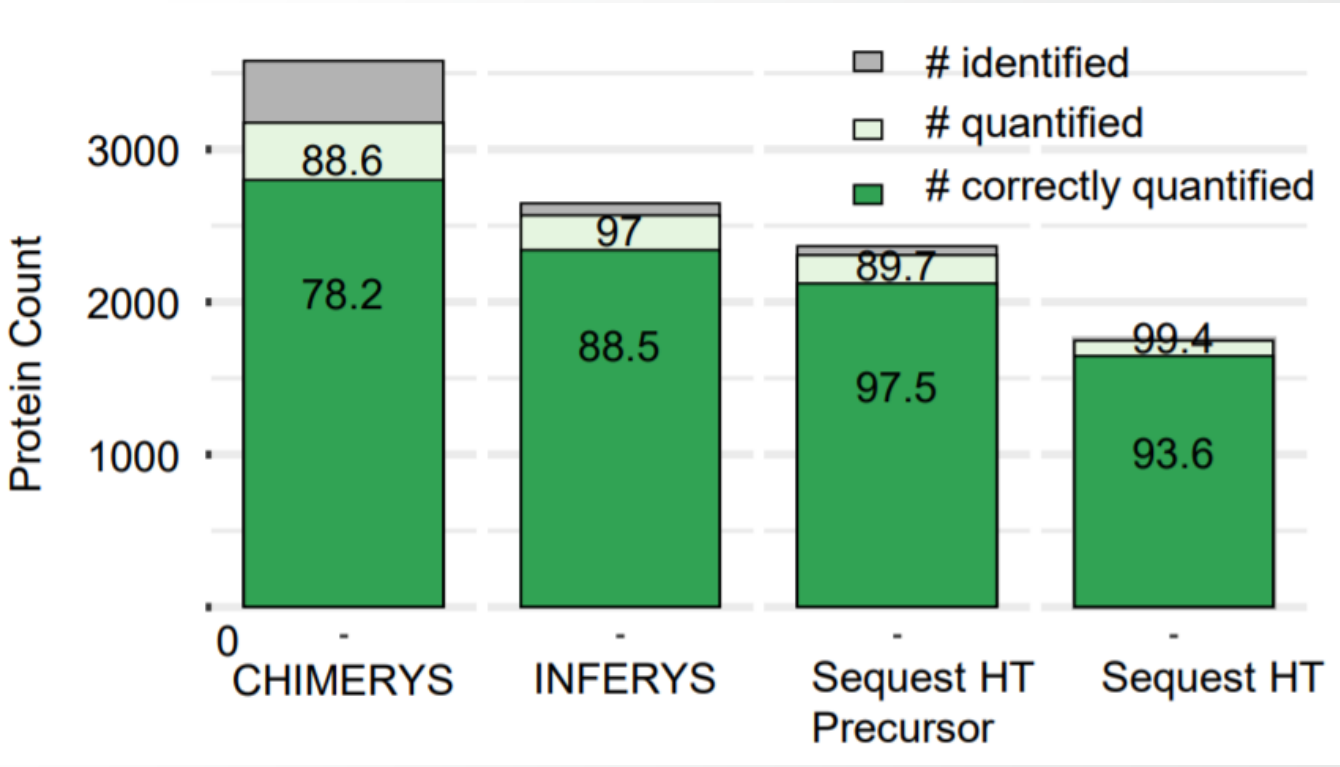
LFQ with CHIMERYS

A two-organism dilution series

- Distribution of quantitative yeast protein ratios from dilution experiment
  (correct: 0.5 * expected < r < 2 * expected)

- CHIMERYS produces more, especially lower abundant quantified peptides and proteins

- 75% more correctly quantified proteins compared to Sequest HT



CHIMERYS - Summary

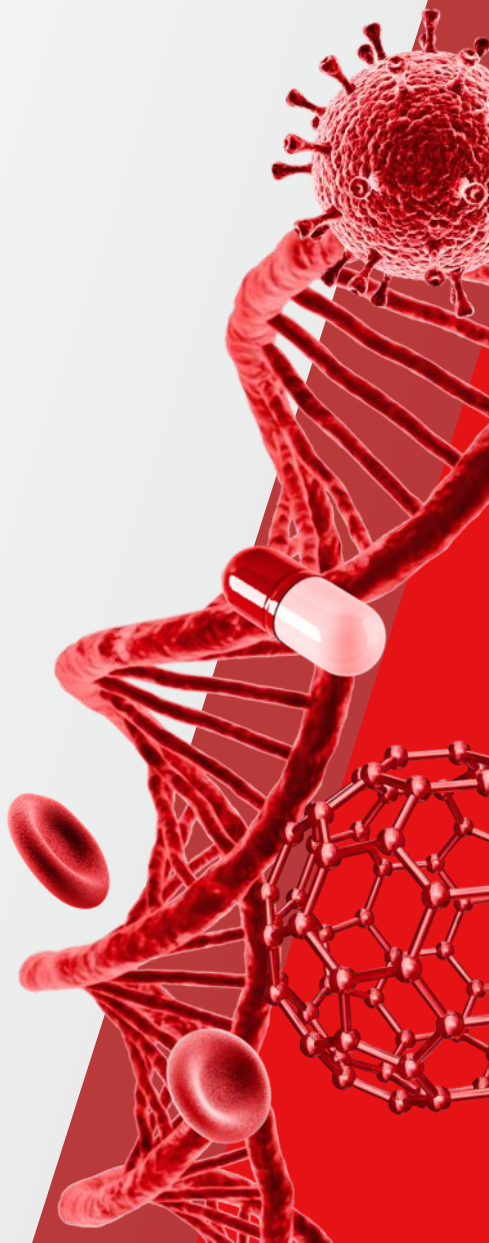
Deconvolutes chimeric spectra based on predicted fragment ion intensities.

- Cloud-native search algorithm that uses AI-based predictions to deconvolute chimeric spectra
- Fully integrated into Proteome Discoverer 3.0 software
- CHIMERYS results in drastically increased numbers of PSM, peptide and protein group identifications, higher sequence coverage and more confident quantification
- CHIMERYS excels at analyzing complex samples, enabling more efficient measurements, advanced acquisition settings and shorter gradients
- Enhanced proteomic throughput, productivity and efficiency

MSAID thermo scientific

ThermoFisher SCIENTIFIC

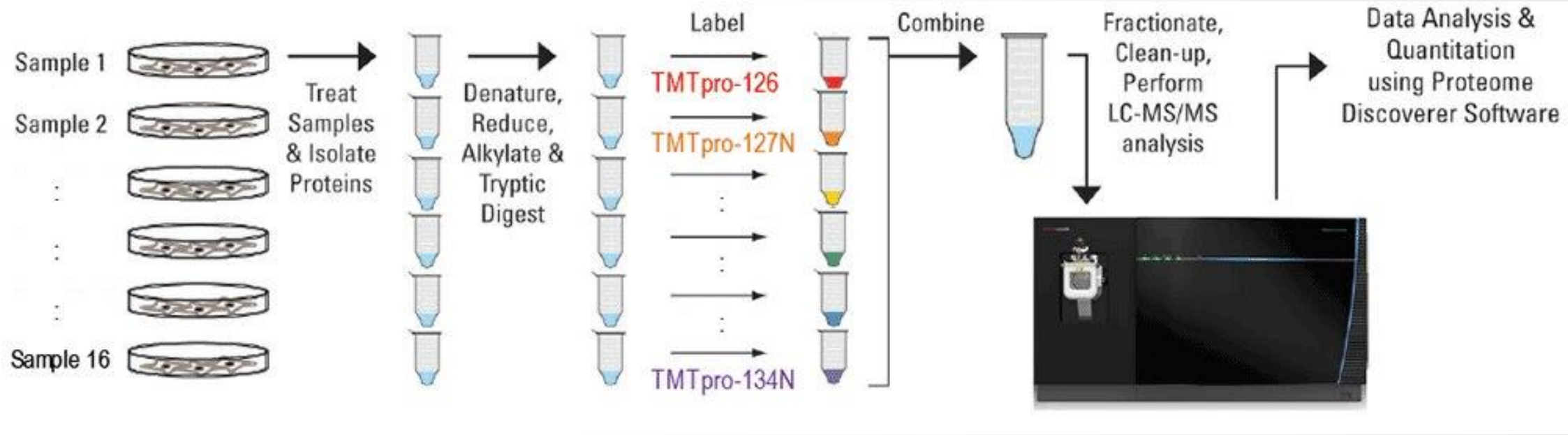
TMTpro 18plex



The world leader in serving science

General Considerations

- Sample complexity
- Number of samples
- Multiplexed experiments
- Instrument acquisition method
- Data processing



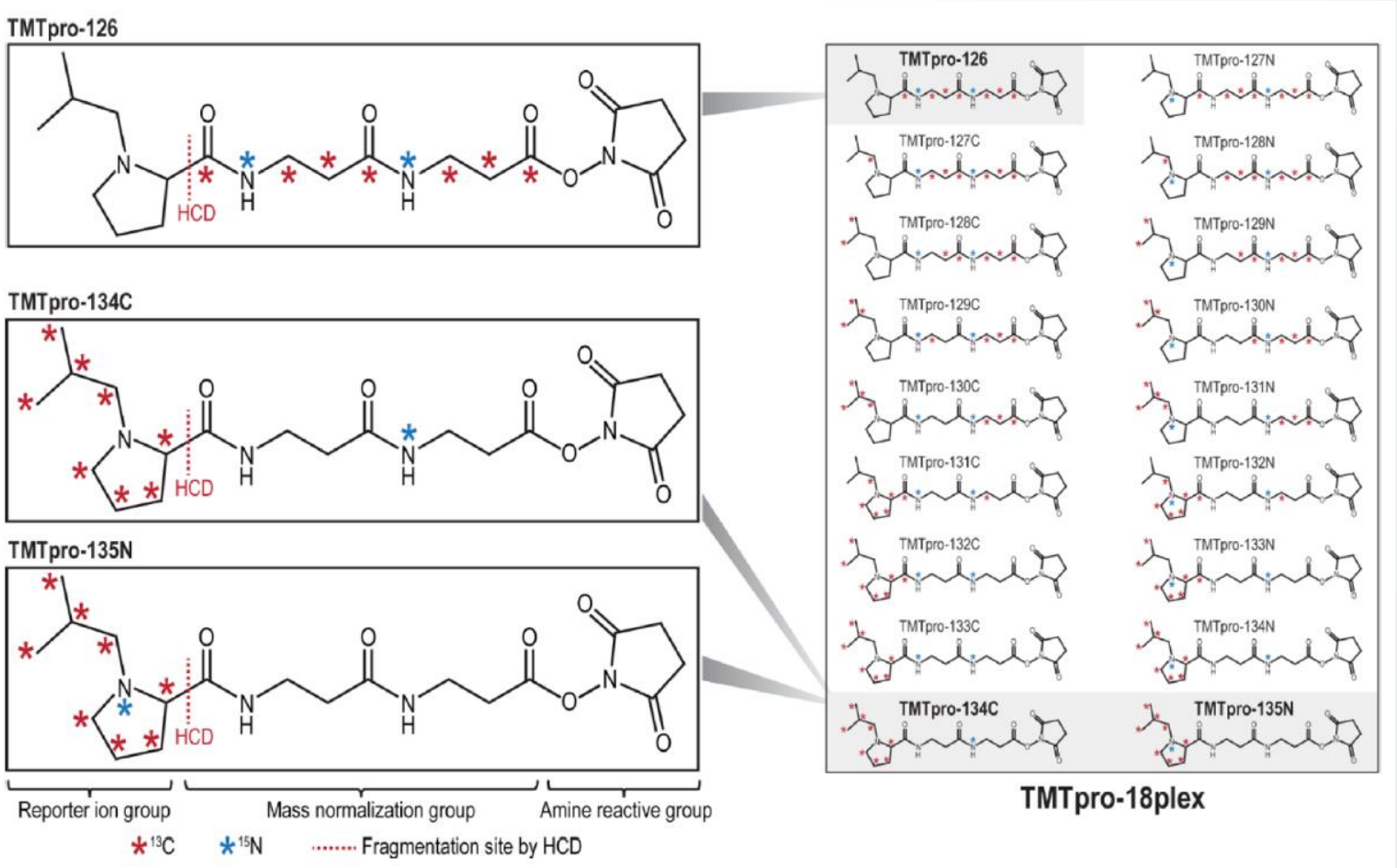
Considerations – Number of samples

- **TMTpro 18plex**
  - Concurrent MS analysis of up to 18 samples derived from cells, tissues, or biological fluids
  - Robust as increased multiplex capability results in fewer missing quantitative values among samples and higher confidence among replicates
  - Efficient amine-reactive, NHS ester-activated reagents ensure efficient labeling of all peptides regardless of protein sequence or proteolytic enzyme specificity
  - Optimized for use with high resolution Thermo Scientific MS/MS platforms

https://doi.org/10.1021/acs.jproteome.1c00168

TMTpro 18plex

- PD 2.5 quan method available
  - https://www.thermofisher.com/order/catalog/product/A52045



Considerations – Acquisition Method Parameters

- TMTpro labeled peptides fragment easier
  - Decrease collision energy
  - MS2: 28-32 NCE
  - MS3: 45-55 NCE
- TMTpro benefits from a better IT scan
  - Use “rapid” scan rather than “turbo”
- TMTpro requires stronger signal
  - 16 (18) vs 11-way split of signal
  - Increase ion fill
  - Load more sample
  - Use wider columns 100-120um
  - Use longer columns (> 35cm)

Properties	Fusion MS2 120 min TMT 11 plex	Fusion MS2 120 min TMT pro 16 plex
Resolution Full MS	120000	120000
AGC target Full MS	4e5	4e5
MS max IT, ms	50	50
Scan range, m/z	400-1400	400-1400
Top Speed, s	3	3
MS2 max IT, ms	120	120
MS2 Isolation window, Th	0.7(2-3)-0.5 (4+)	0.7
MS2 NCE, %	38-40	35
MS2 Intensity threshold	5e4	5e4
Dynamic exclusion, s	60, single charge	60, single charge
MS2 Resolution	50000	50000
MS2 AGC target	1e5	1.2e5
MS3 AGC target		
SPS Isolation window, Th		
SPS NCE, %		
SPS max IT, ms		
SPS settings: # notches, mass range, Tag Exclusion	*m/z* 110	*m/z* 110

Considerations – Acquisition Strategy

- FAIMS
  - Can decrease co-isolation interference (ratio compression)
- MS2 or MS3 strategy
  - Is ratio accuracy that important?
- Real Time Search
  - Available on Eclipse
- Close Out function
  - Specify the maximum number of peptides per protein to be quantified
- Turbo/Rapid scan
  - Higher number of IDs in MS2 run
  - “Rapid” recommended for TMTpro

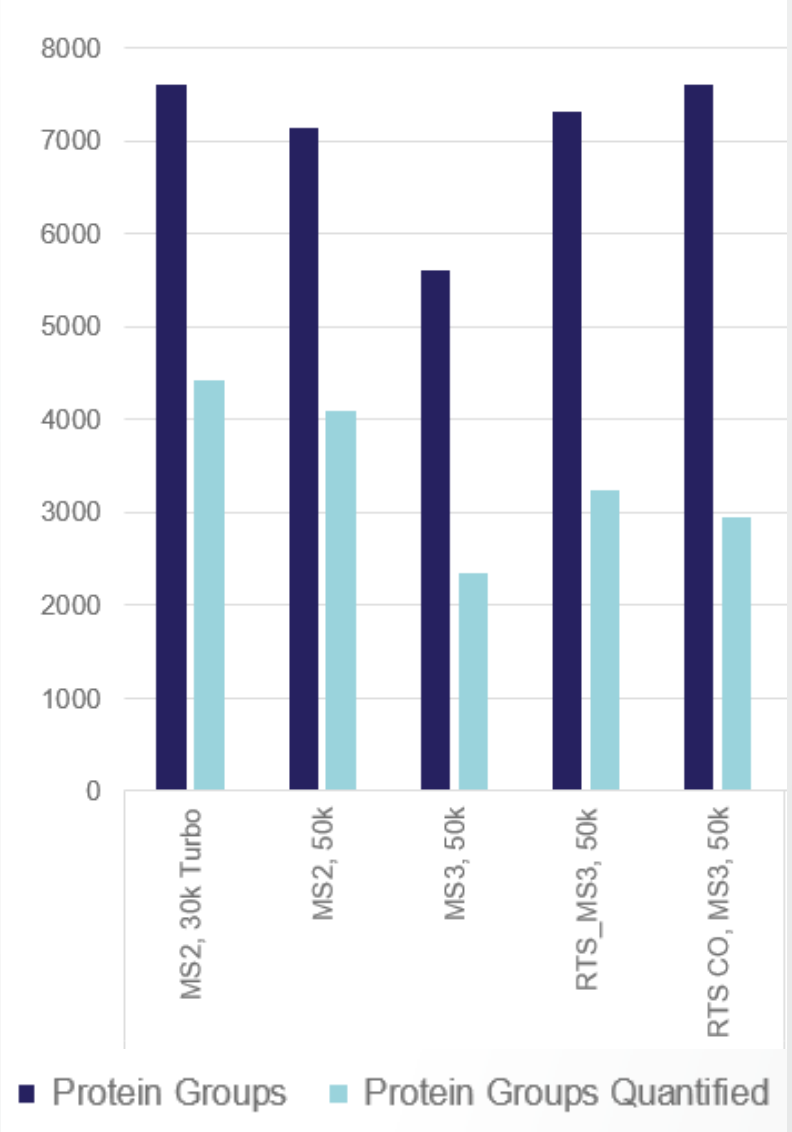
Considerations – Acquisition Strategy

Complex mixture, 50 cm column, 90 min gradient, 6 fractions

- Which acquisition strategy is most appropriate?

- MS2 30k Turbo scan
- MS2 50k
- MS3 50k
- RTS, MS3 50k
- RTC CO, MS3 50k

- Caveat
  - Protein ID being the only figure of merit
  - Quan accuracy and precision not considered

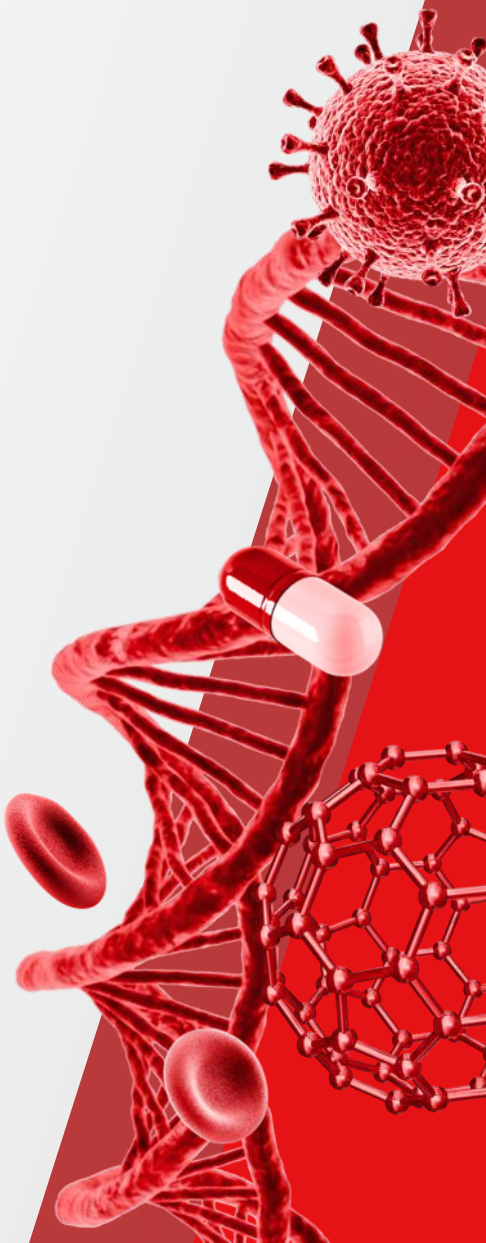


Data courtesy of Jenny Ho (TFS Hemel Hempstead)

ThermoFisher SCIENTIFIC

Reporter Ions Quantifier node

PSM Filters for quantification

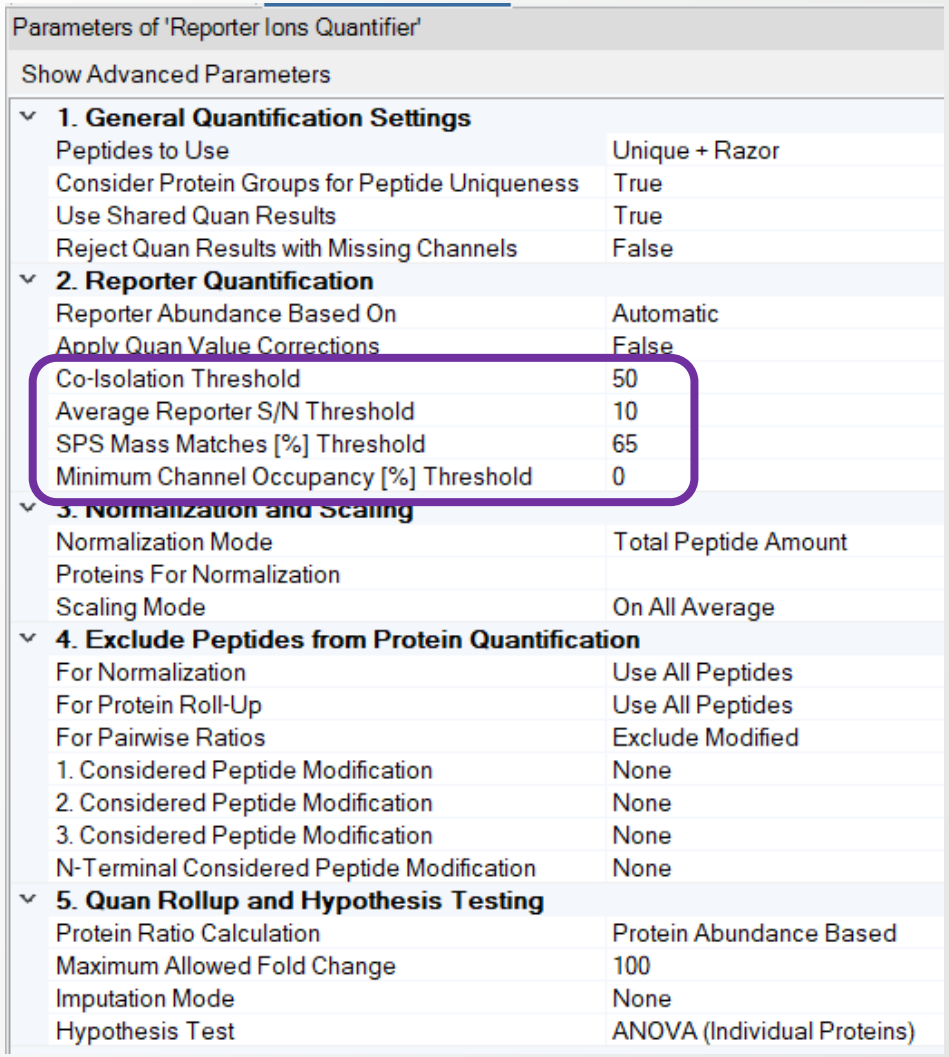


The world leader in serving science

Reporter Ions Quantifier node

Discussion of parameter settings

- Co-isolation threshold – tries to minimize ratio compression due to chimeric spectra
- Average reporter S/N threshold – avoids issues due to ion statistics
- SPS mass matches threshold – considers number of relevant ions selected from MS2
- Minimum channel occupancy - only peptides with at least this percentage of non-zero channel values would be considered for quantification



Reporter Ions Quantifier - Filters

- **Co-isolation threshold**

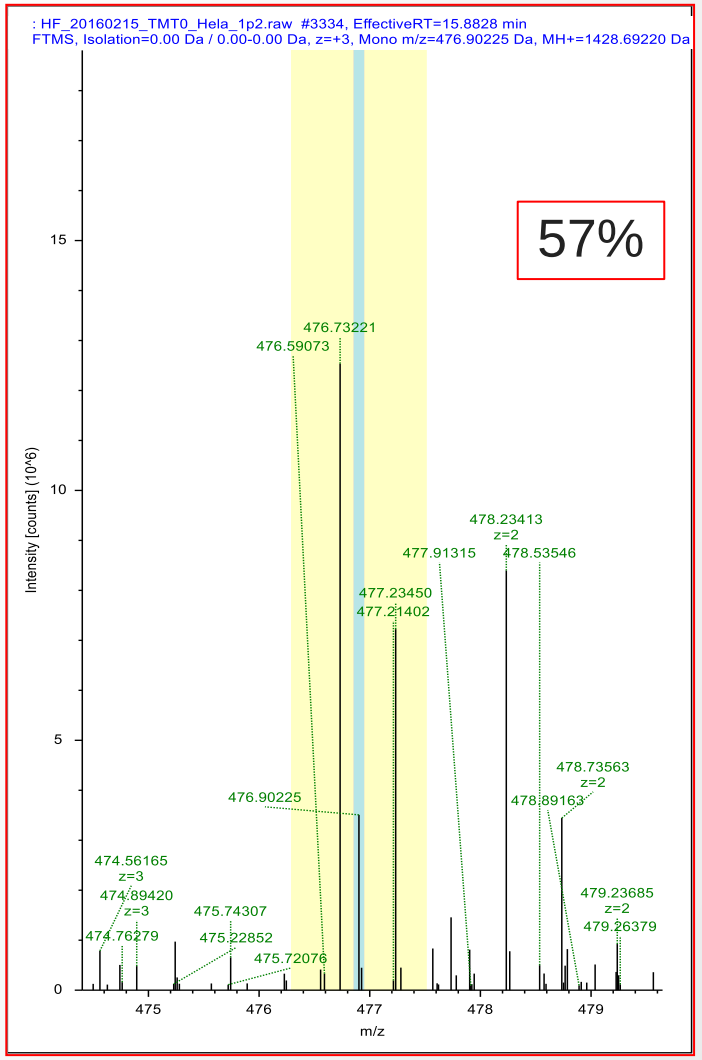
$$\%\_isolation\_interference = 100 \times \left[1 - \left(\frac{precursor\_intensity\_in\_isolation\_window}{total\_intensity\_in\_isolation\_window}\right)\right]$$

- If the contribution of unwanted peaks in the MS1 isolation window exceeds the threshold, this filter will exclude MS2 or a dependent MS3 spectrum from contributing to quantification

Reporter Ions Quantifier - Filters

- **Co-isolation threshold**
- This PSM would be *“excluded by method”*

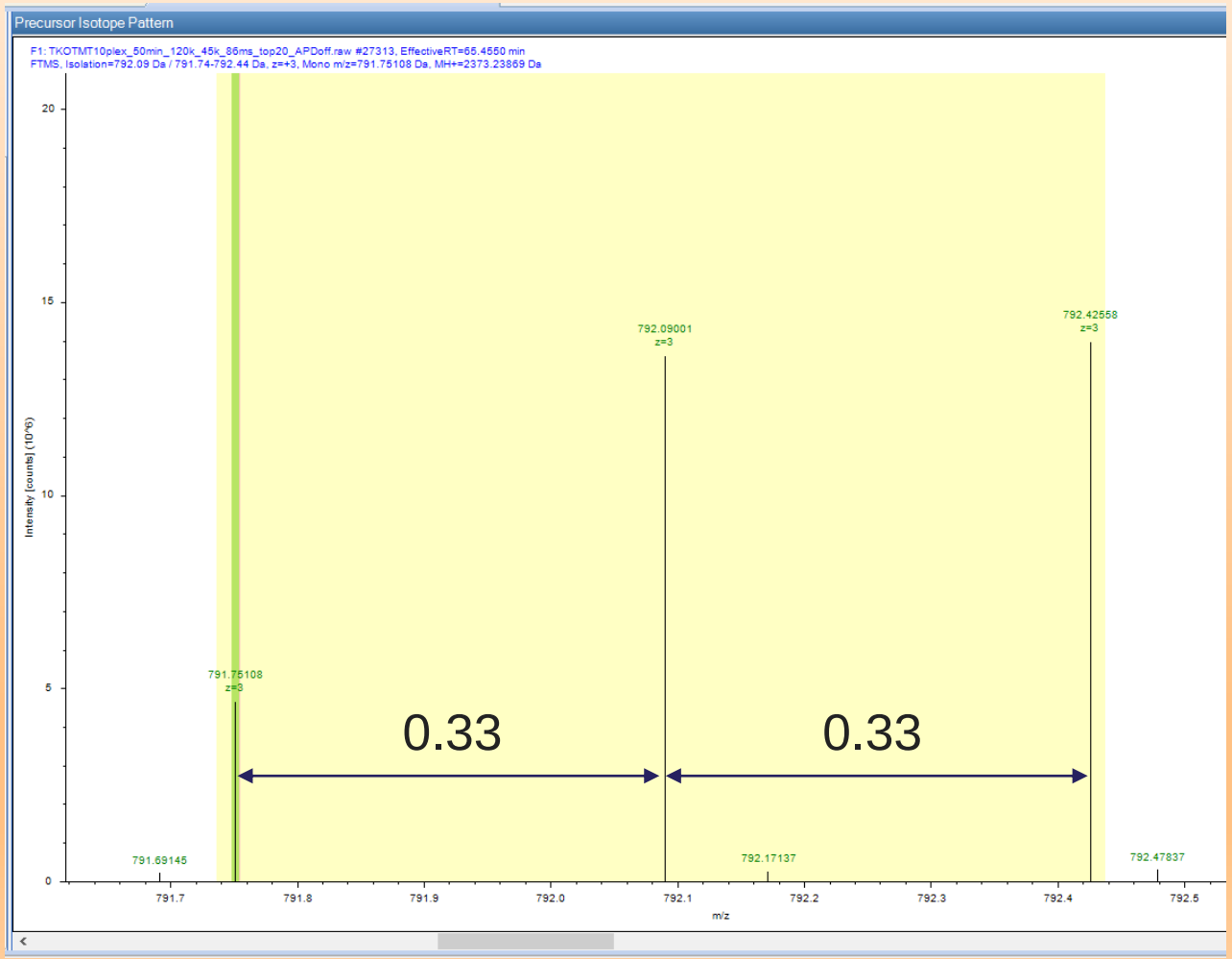
Parameters of 'Reporter Ions Quantifier'	
Show Advanced Parameters	
1. General Quantification Settings	
Peptides to Use	Unique
Consider Protein Groups for Peptide Uniqueness	True
Use Shared Quan Results	True
Reject Quan Results with Missing Channels	False
2. Reporter Quantification	
Reporter Abundance Based On	Automatic
Apply Quan Value Corrections	True
Co-Isolation Threshold	50
Average Reporter S/N Threshold	10
SPS Mass Matches [%] Threshold	65



Q: Is there a bug in co-isolation threshold calculation?

- Isotopes of the selected parent ion *sometimes* considered “contaminants”
  - Example 3+ peptide dYGWTQTSLDDYPk
  - isolation interference reported = 86%

- Use isolation width <0.65 Da



Reporter Ions Quantifier - Filters

- **Average Reporter S/N Threshold**
- “Magic” number = 10
- Sum of S/N for all reporter ions in the spectrum = “10” x plex of the method

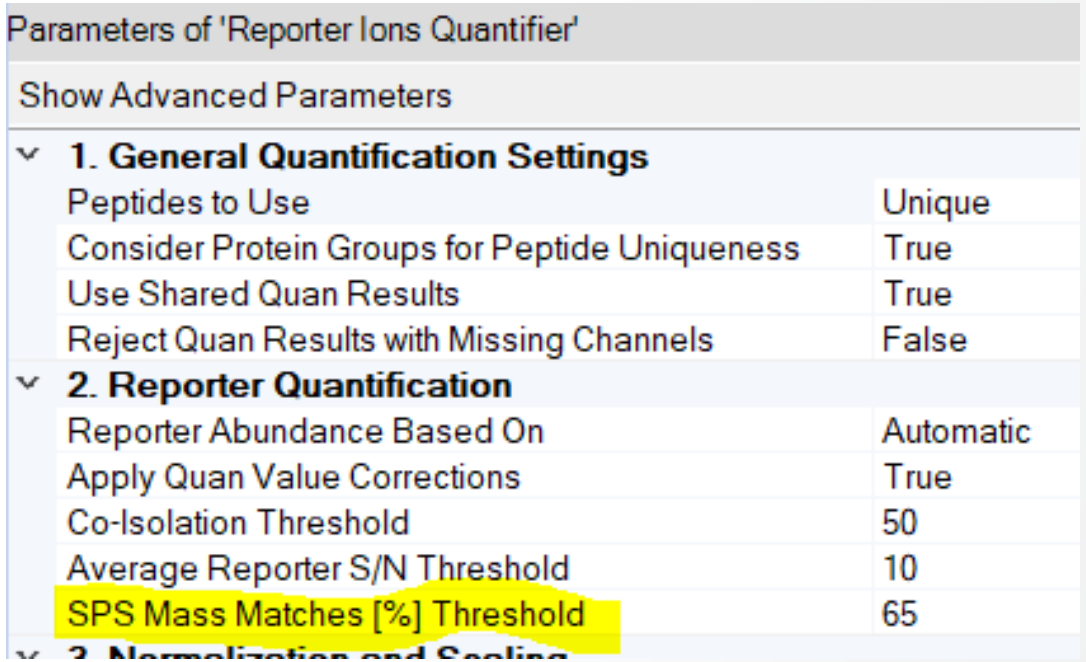
Parameters of 'Reporter Ions Quantifier'	
Show Advanced Parameters	
1. General Quantification Settings	
Peptides to Use	Unique
Consider Protein Groups for Peptide Uniqueness	True
Use Shared Quan Results	True
Reject Quan Results with Missing Channels	False
2. Reporter Quantification	
Reporter Abundance Based On	Automatic
Apply Quan Value Corrections	True
Co-Isolation Threshold	50
Average Reporter S/N Threshold	10
SPS Mass Matches [%] Threshold	65

Example TMT 6plex
Sum of reporters S/N > 10 x 6

Example TMT 11plex
Sum of reporters S/N > 10 x 11

Reporter Ions Quantifier - Filters

- **SPS Mass Matches [%] Threshold**
  - *Only for MS3-based quan*
  - Min. percentage of MS2 fragments selected for MS3 fragmentation attributable to the precursor ion



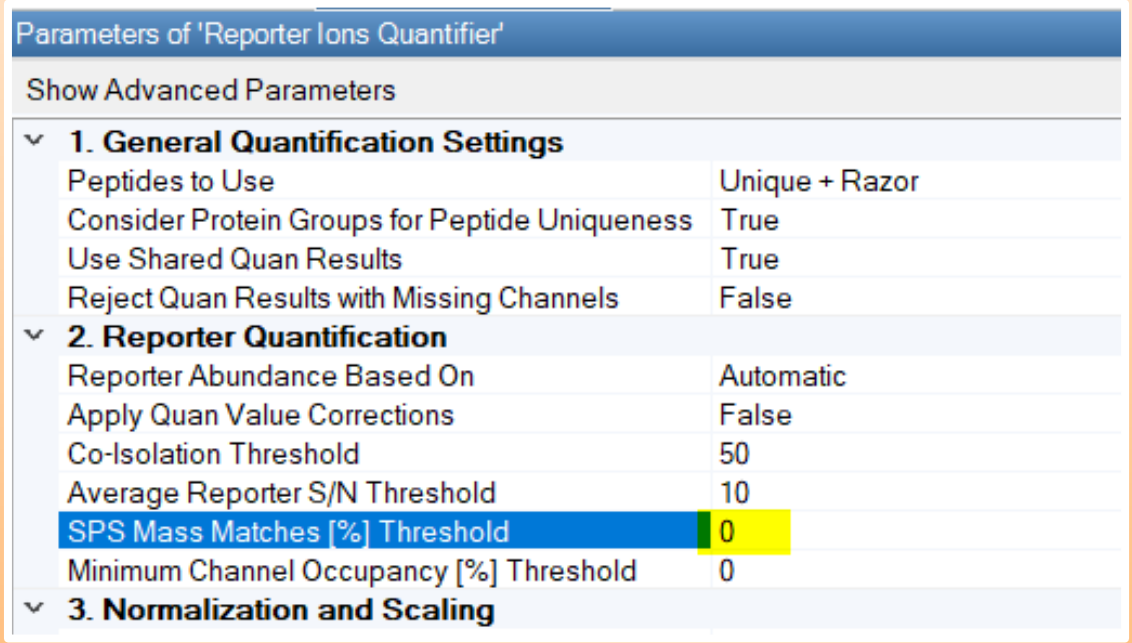
Q: Is there a bug in SPS Mass Matches Code?

ThermoFisher SCIENTIFIC

TMT-labeled phosphopeptides

- For SPS MS3 data → SPS Mass Matches [%] Threshold = 0

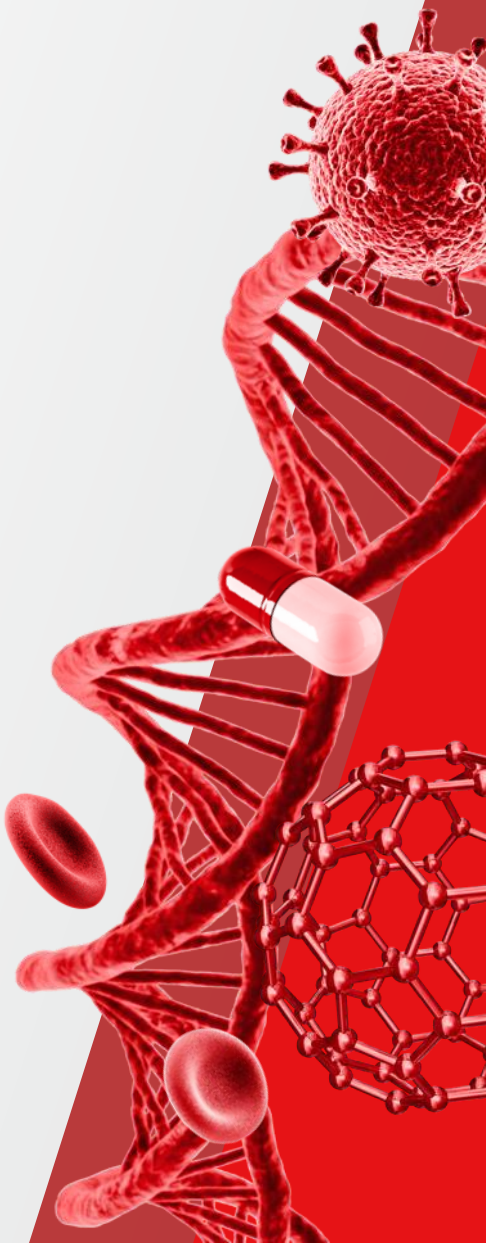
- SPS Mass Match [%] Threshold considers MS1 precursor mass to check whether MS2 fragments belong to the peptide of interest
- The precursor carries the phosphate but most MS2 fragments are likely to contain dephosphorylated residues (dehydroalanine instead of phosphoserine)
- PD concludes that such fragments are not from the peptide of interest → Phosphopeptide MS3 spectra are unlikely to pass the filter…



ThermoFisher SCIENTIFIC

Reporter Ions Quantifier node

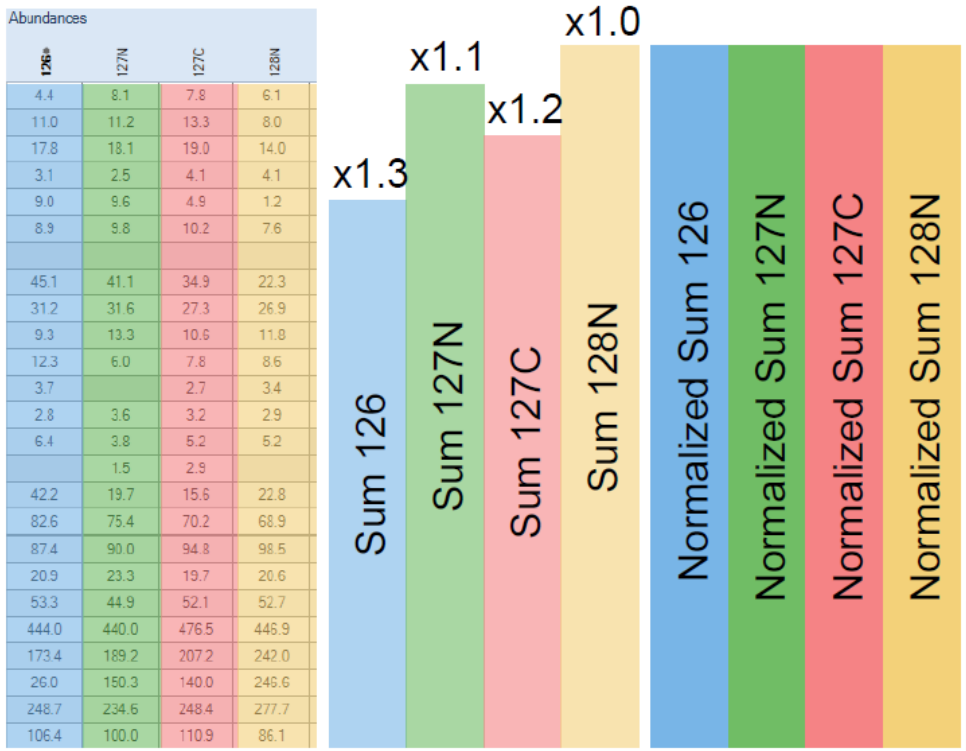
Normalization and Scaling



The world leader in serving science

Normalization

- Total peptide amount
  - Sum reporter ions per channel
  - Calculate normalization factor



Normalization Mode
Specifies how the normalization shall be performed to correct for experimental bias.

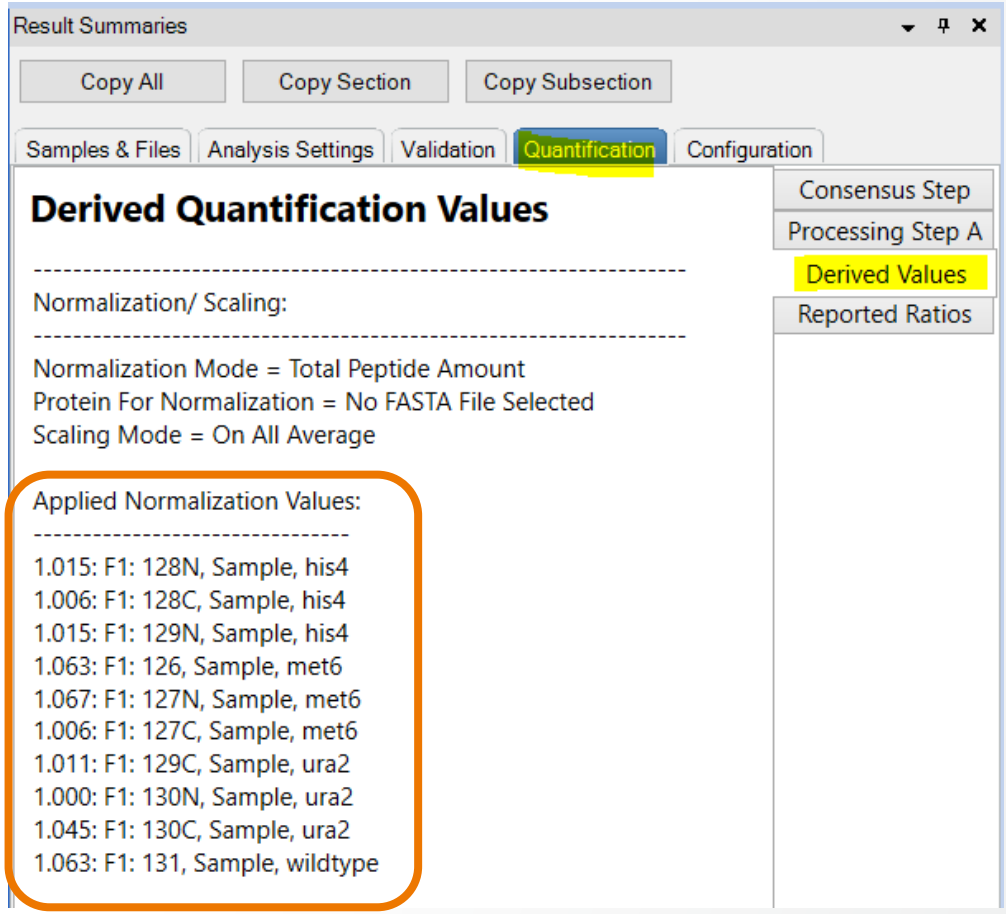
None: No normalization is applied.

Total Peptide Amount: This calculates the total sum of the abundance values for each channel over all peptides identified within a file. It then takes the channel with the highest total abundance as a reference and corrects all abundance values in all other channels by a constant factor per channel, so that at the end the total abundance is the same for all channels.

Specific Protein Amount: In this mode, the normalization only looks at the summed abundances of proteins in the specified FASTA file. This is useful if you know there are housekeeping proteins, or a bait protein from pull-downs, or similar proteins for which you can assume their abundance remaining constant across your treatments. Besides that the normalization is done similar to the normalization to total peptide amount.

Normalization

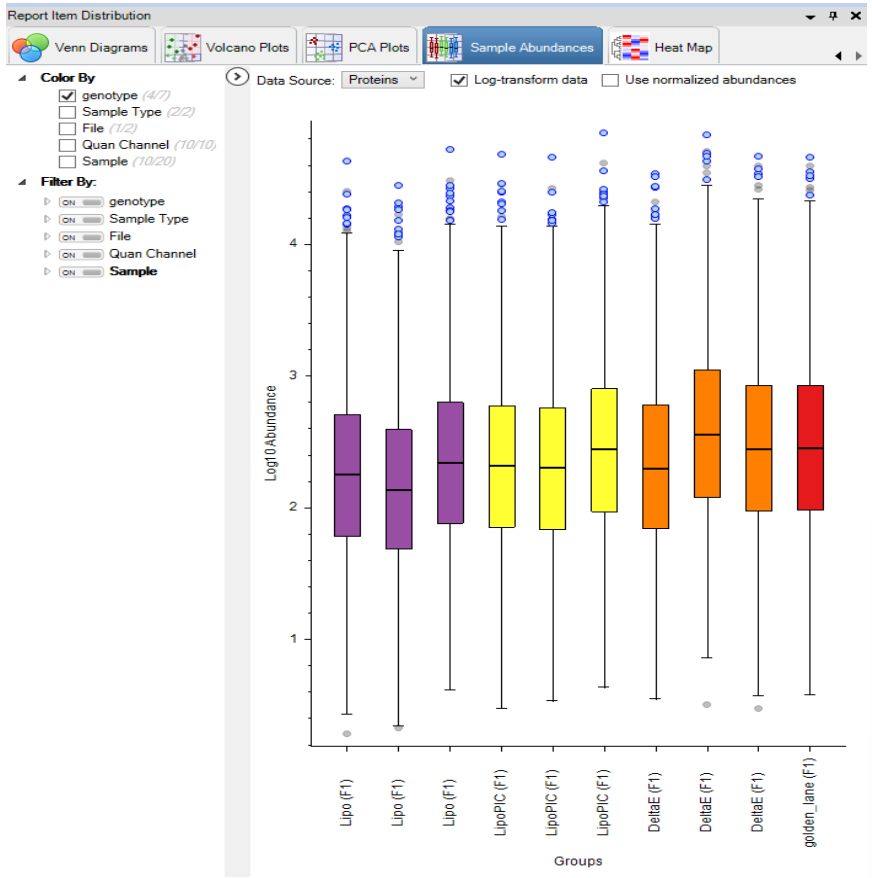
- Applied normalization values
- Result Summaries → Quantification → Derived Values



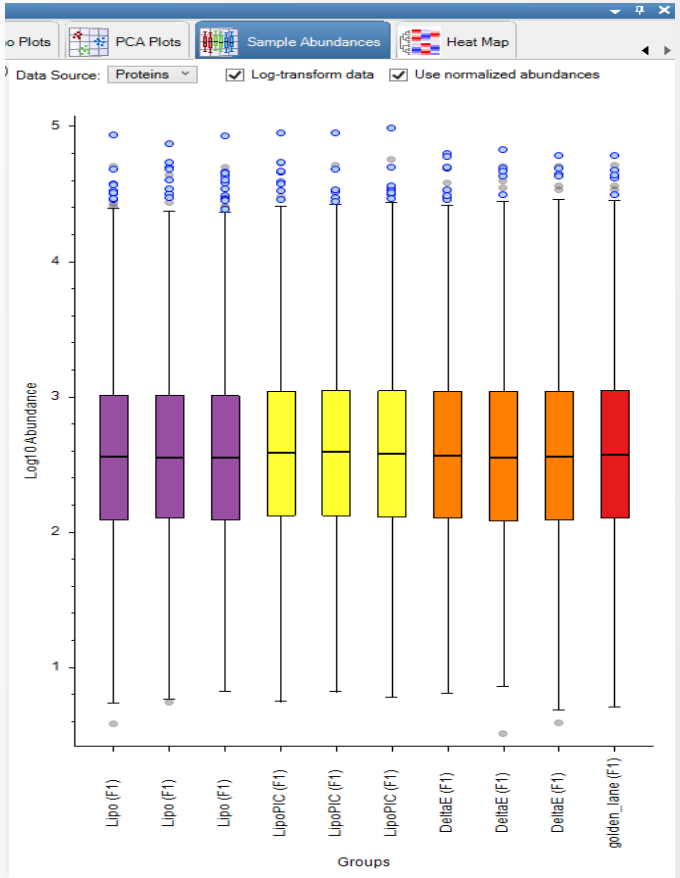
Normalization

- Check how normalization worked
- View → Distribution Charts → Sample Abundances

BEFORE

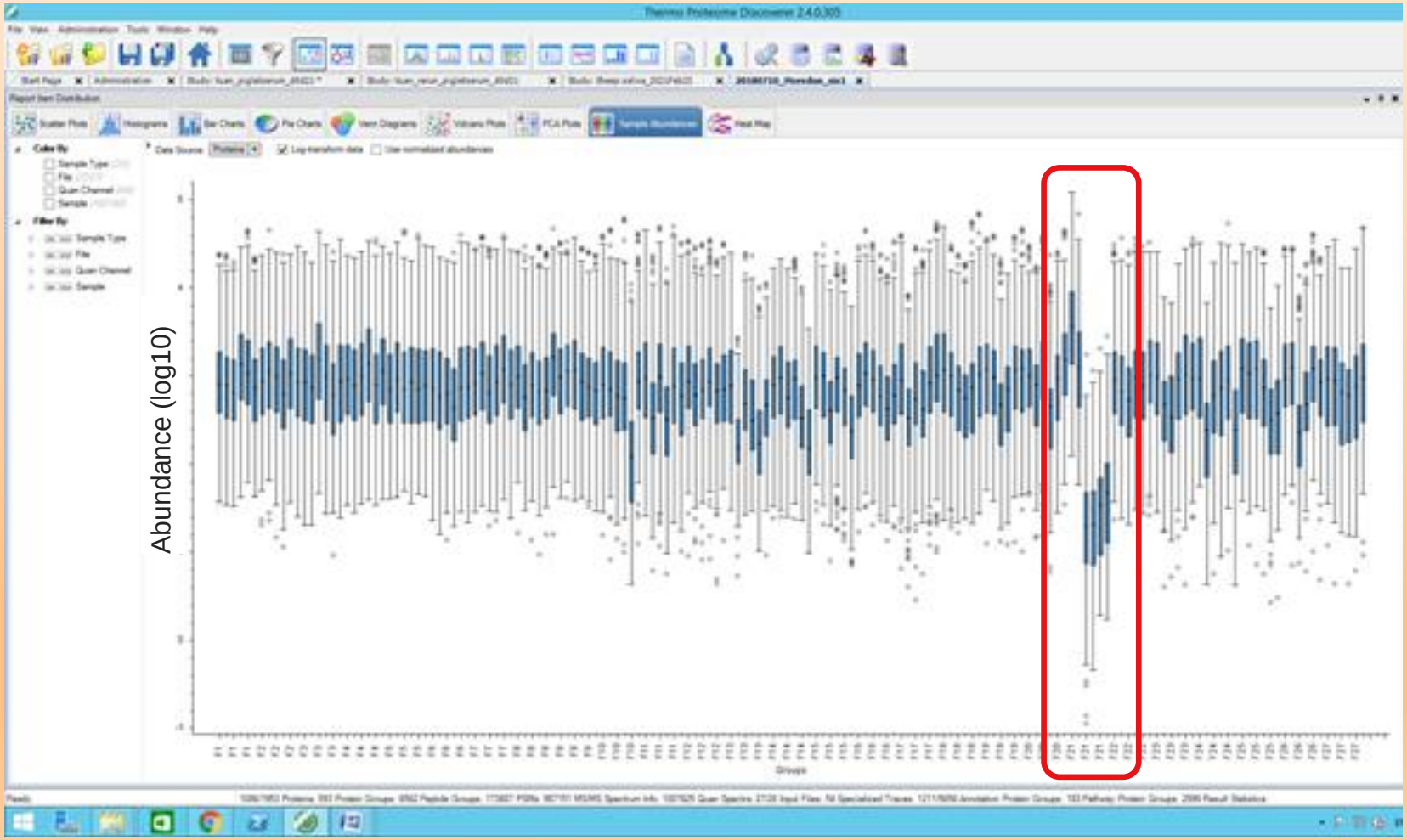


AFTER



Q: What went wrong ?

- View → Distribution Charts → Sample Abundances
- Only 2-3 fold difference in sample load acceptable



Multiplexed TMT Experiment

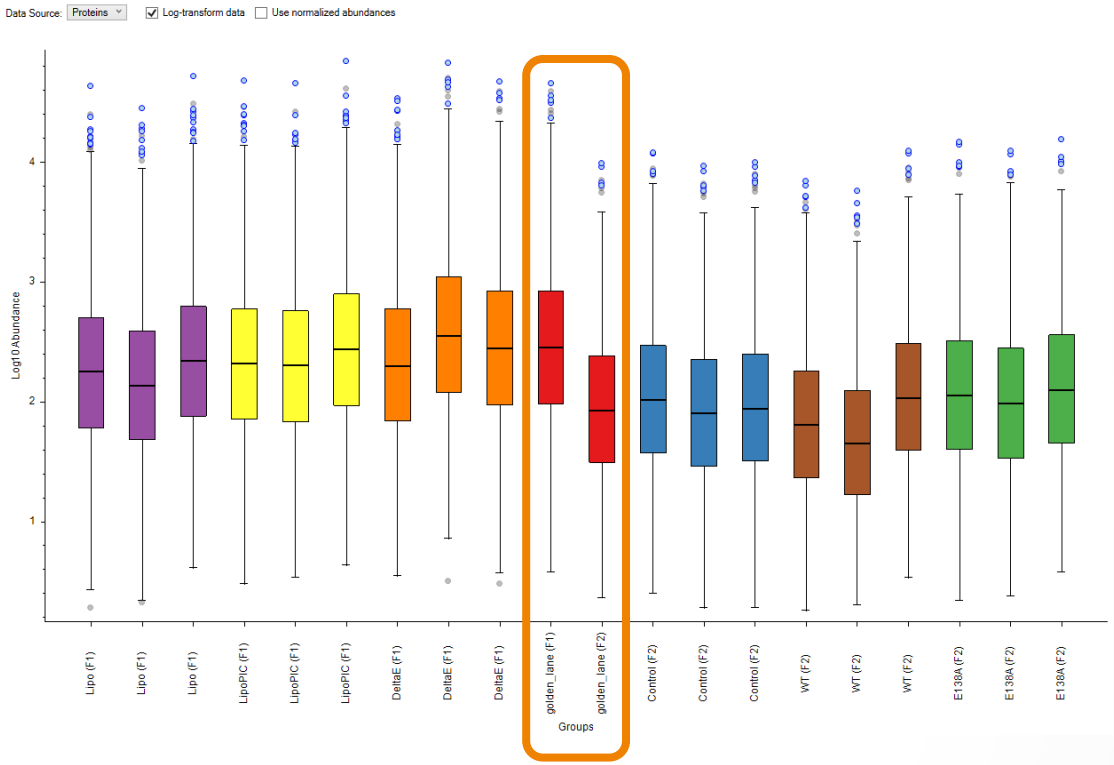
- The number of samples exceeds the number of available channels
  - The same label will be used for two or more different samples
  - Those will be analyzed in separate “sets”

- *Example of a multiplexed experiment:*
  - You have 6 conditions with 3 samples each (total 18 samples)
  - You have TMT 10plex kit available
  - Split sample to two batches; 9 samples each
  - Prepare a bridge sample (e.g., “pooled” sample, run as the 10th channel)
  - Run 2 TMT 10plex sets
  - *Both runs should be using the same quan method!*

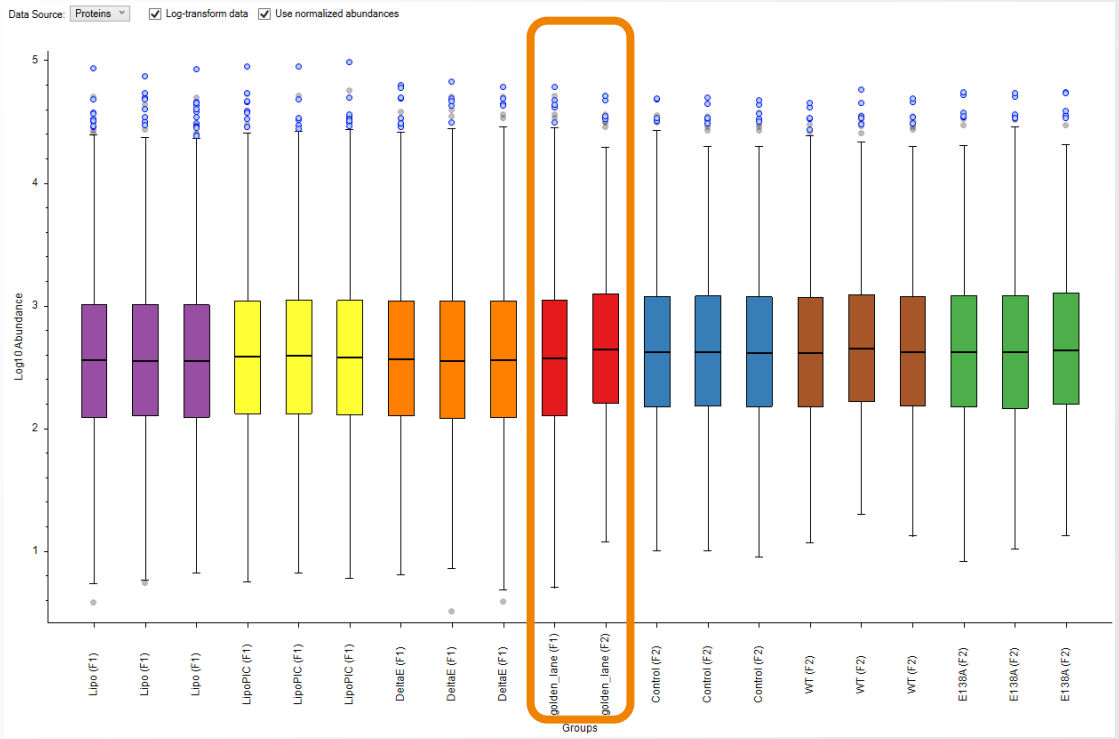
Normalization of Multiplexed TMT Experiment

- 2 sets of TMT 10plex

BEFORE

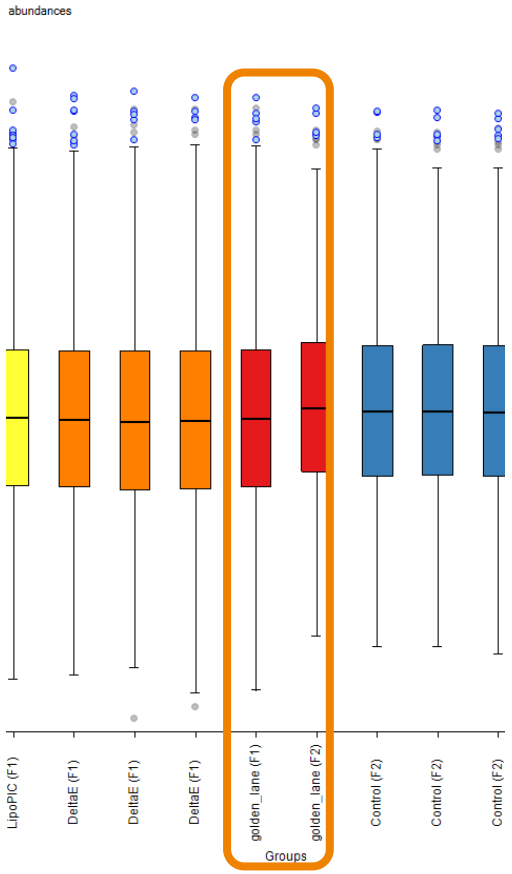


AFTER



Normalization of Multiplexed TMT Experiment

- *Note:* the median abundance of the two bridge samples is not identical
- The normalization works only within the same set, not across different sets



Scaling

Use the bridge channels to normalize across multiple sets

- Set Scaling On **controls** average

Parameters of 'Reporter Ions Quantifier'

Hide Advanced Parameters

1. General Quantification Settings	
Peptides to Use	Unique + Razor
Consider Protein Groups for Peptide Unique	True
Use Shared Quan Results	True
Reject Quan Results with Missing Channels	False
2. Reporter Quantification	
Reporter Abundance Based On	Automatic
Apply Quan Value Corrections	False
Co-Isolation Threshold	50
Average Reporter S/N Threshold	10
SPS Mass Matches [%] Threshold	65
Minimum Channel Occupancy [%] Threshol	0
3. Normalization and Scaling	
Normalization Mode	Total Peptide Amount
Proteins For Normalization	
Scaling Mode	On Controls Average

Define the “controls” in Samples table

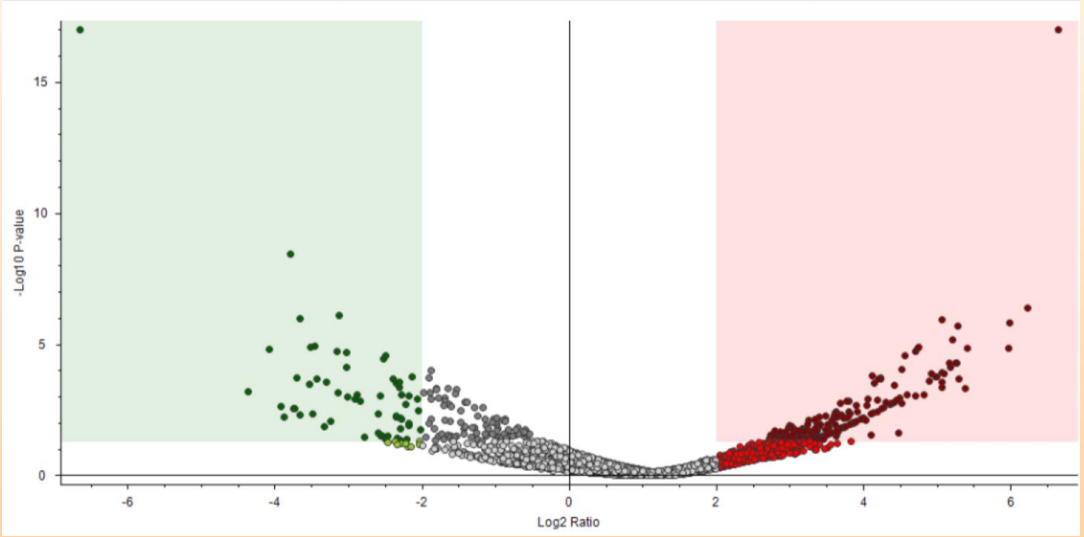
Add Files | Add Fractions | Remove Files | Open Containing Folder | New Analysis

Study Definition | Input Files | Samples | Analysis Results

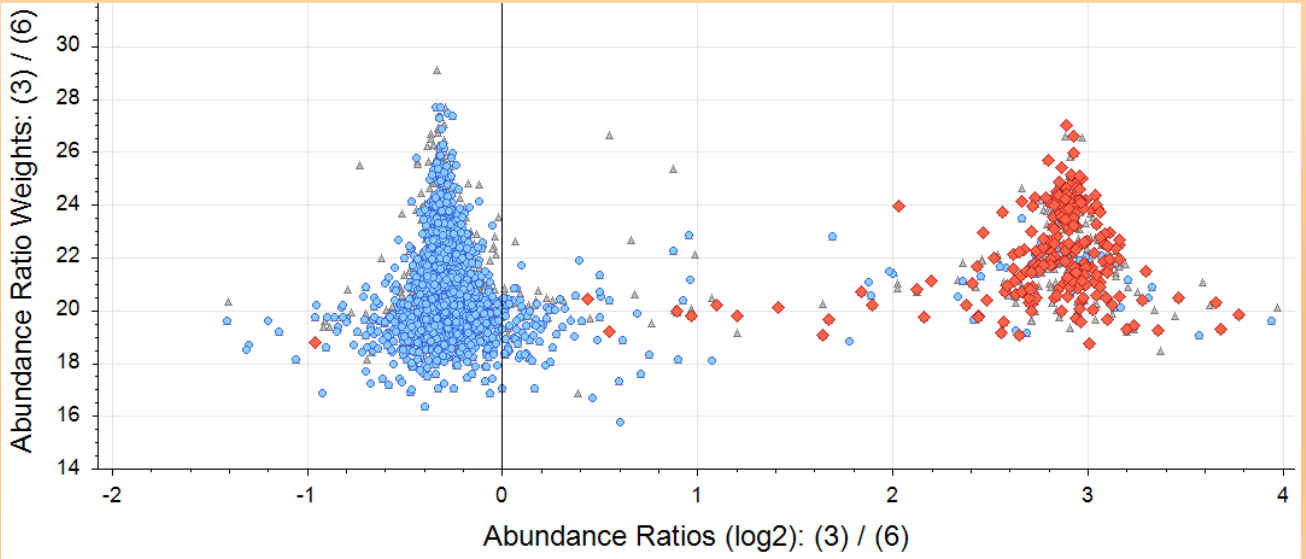
Error	Sample	File	Sample Identifier	Sample Type	Quan Chani
	S1	F1	BDL_TMT_SLD_MY_TD_01 - [113]	Control	113
	S2	F2	BDL_TMT_SLD_MY_TD_02 - [113]	Control	113
	S3	F3	BDL_TMT_SLD_MY_TD_03 - [113]	Control	113
	S4	F4	BDL_TMT_SLD_MY_TD_04 - [113]	Control	113
	S5	F1	BDL_TMT_SLD_MY_TD_01 - [114]	Sample	114
	S12	F2	BDL_TMT_SLD_MY_TD_02 - [114]	Sample	114
	S19	F3	BDL_TMT_SLD_MY_TD_03 - [114]	Sample	114
	S26	F4	BDL_TMT_SLD_MY_TD_04 - [114]	Sample	114
	S6	F1	BDL_TMT_SLD_MY_TD_01 - [115]	Sample	115
	S13	F2	BDL_TMT_SLD_MY_TD_02 - [115]	Sample	115
	S20	F3	BDL_TMT_SLD_MY_TD_03 - [115]	Sample	115
	S27	F4	BDL_TMT_SLD_MY_TD_04 - [115]	Sample	115

Q: Have I selected a correct normalization method ?

Check the volcano plot

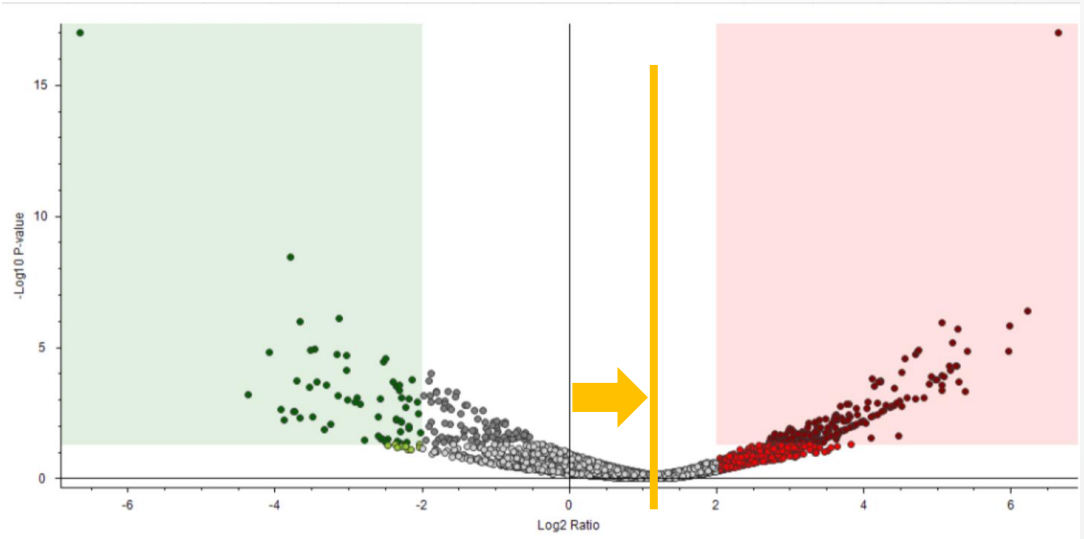


or even better, the plot of
Abundance Ratio Weights x
Abundance ratios (log2)



Q: Have I selected a correct normalization method ?

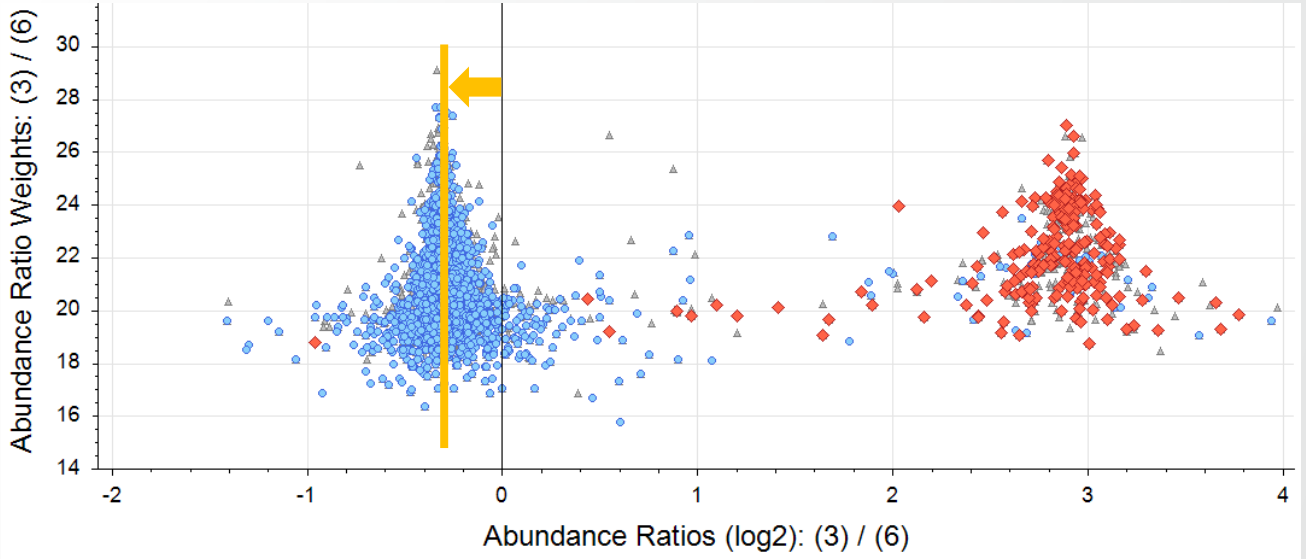
Check the volcano plot



or even better, the plot of

Abundance Ratio Weights x

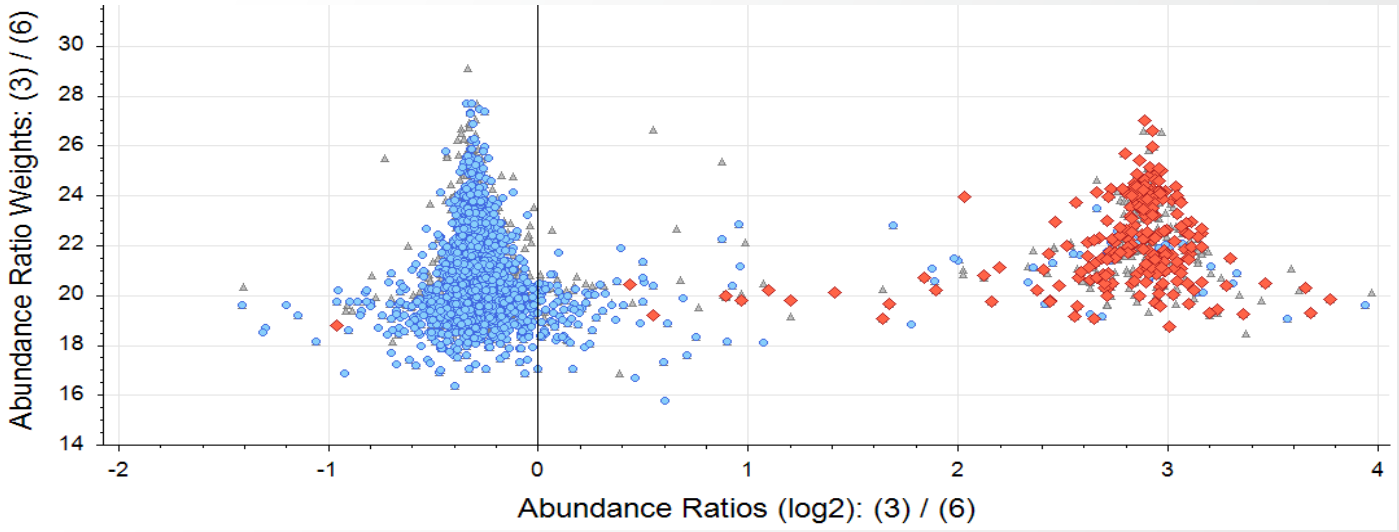
Abundance ratios (log2)



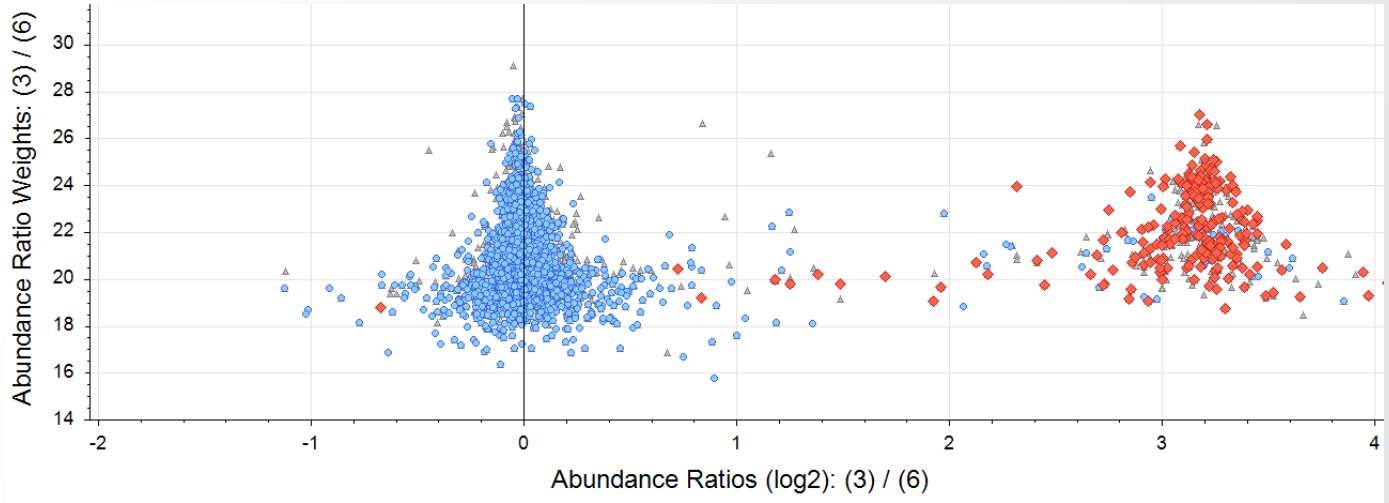
Different Proteins Used for Normalization

Mixed proteomes dilution experiment: identical human background; different bacterial spike-in

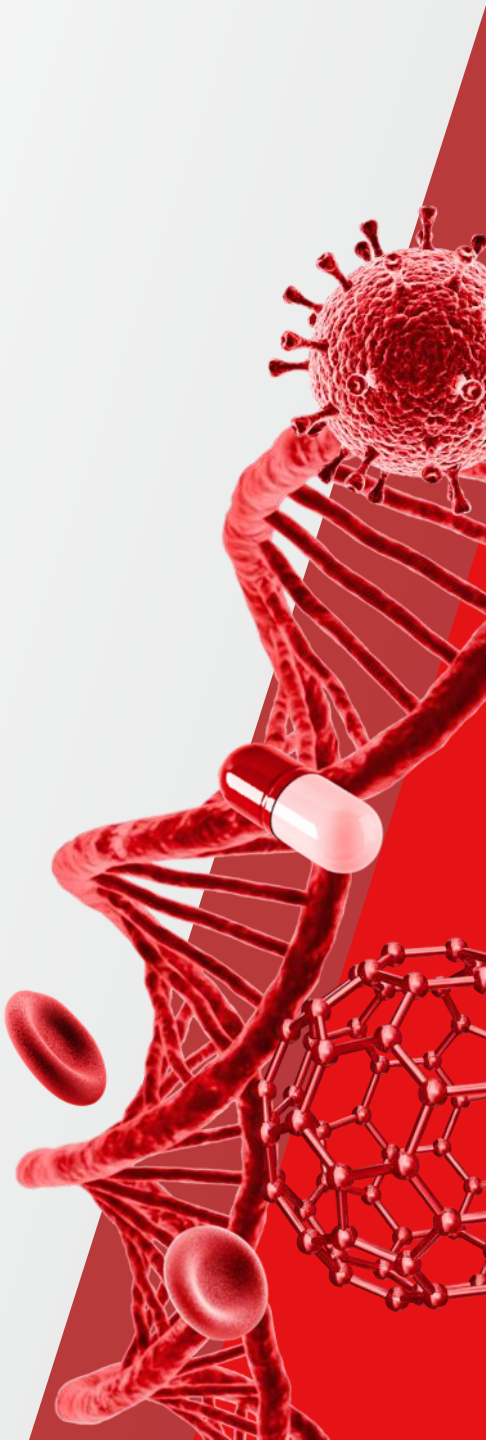
- Normalized on total peptide amount
  - The background protein distribution (blue = human) is not centered at log2=0
  - The total protein amount in the two samples is different



- Normalized on specific protein amount
  - "human.FASTA" db used
  - Normalized on all human peptides (amount in both samples identical)



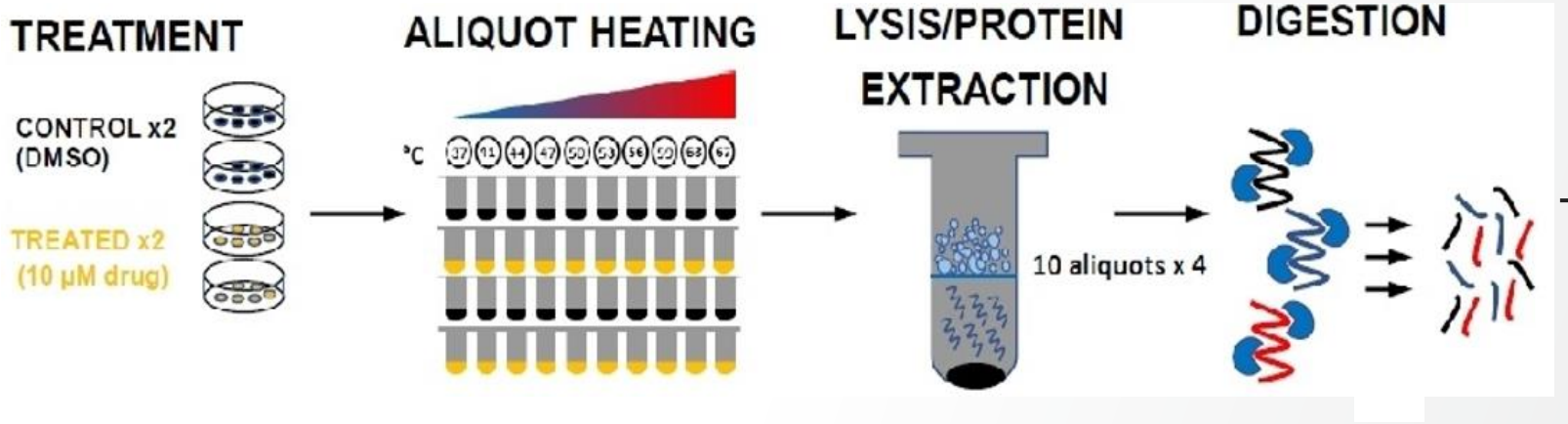
Multiplexed TMT - Experiment Setup



The world leader in serving science

Multiplexed TMT11plex Experiment

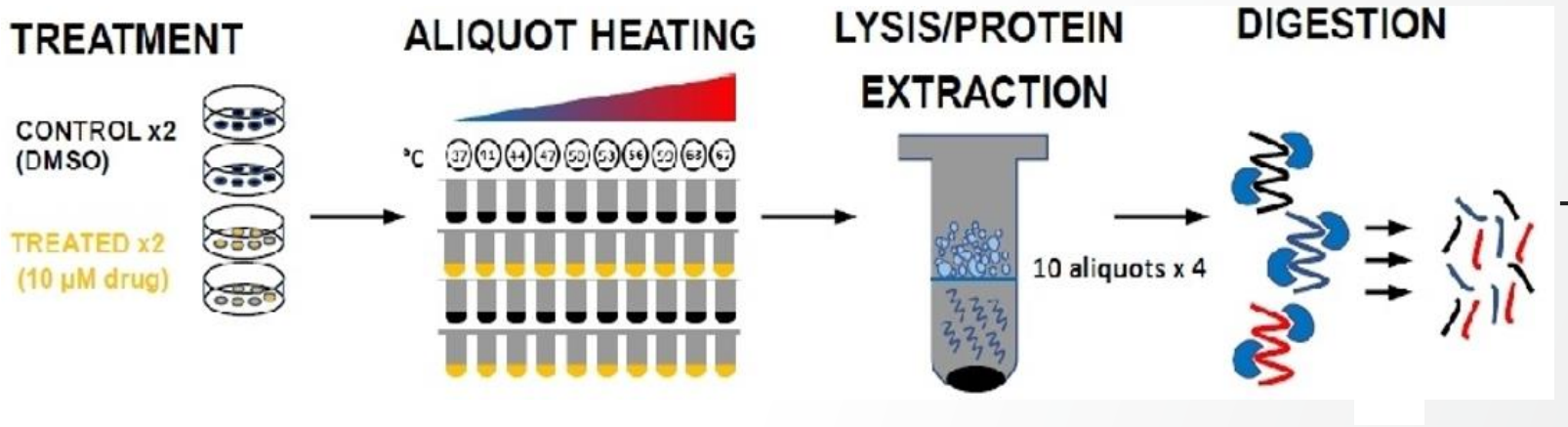
- 2 treatments (DMSO; drug)
- 2 biological replicates
- 10 temperature points (channels 126-131N)
- 1 “pool” sample (bridge channel 131C)



T (°C)	Channel
37	126
41	127N
44	127C
47	128N
50	128C
53	129N
56	129C
59	130N
63	130C
67	131N
Pool	131C

Multiplexed TMT11plex Experiment

- To accommodate all samples, we need **4x TMT11plex sets**
- Each set **fractionated** (2 fractions each)
- Each fraction **run in triplicate** **=> Total 24 .raw files**

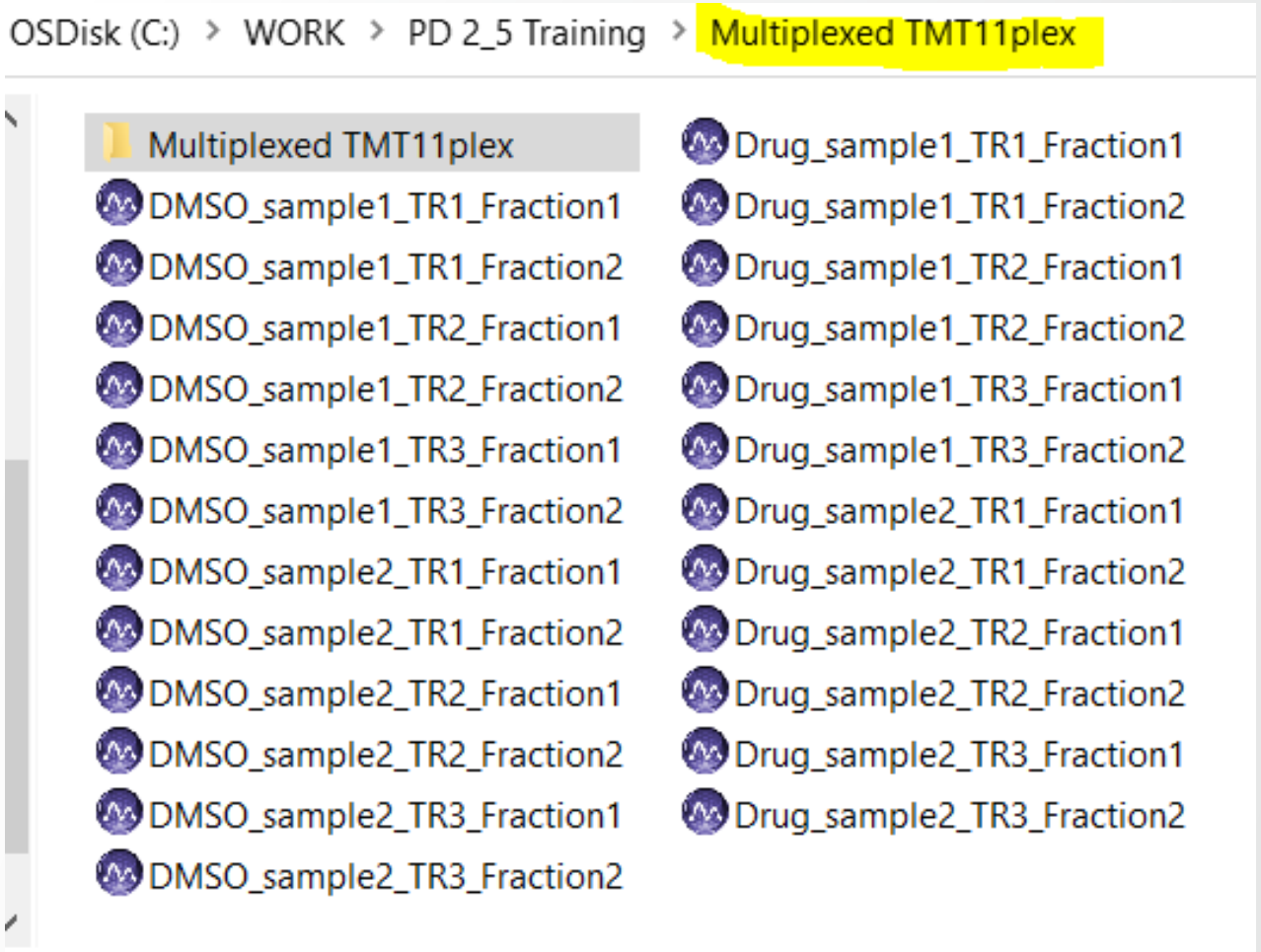


T (°C)	Channel
37	126
41	127N
44	127C
47	128N
50	128C
53	129N
56	129C
59	130N
63	130C
67	131N
Pool	131C

Multiplexed TMT11plex

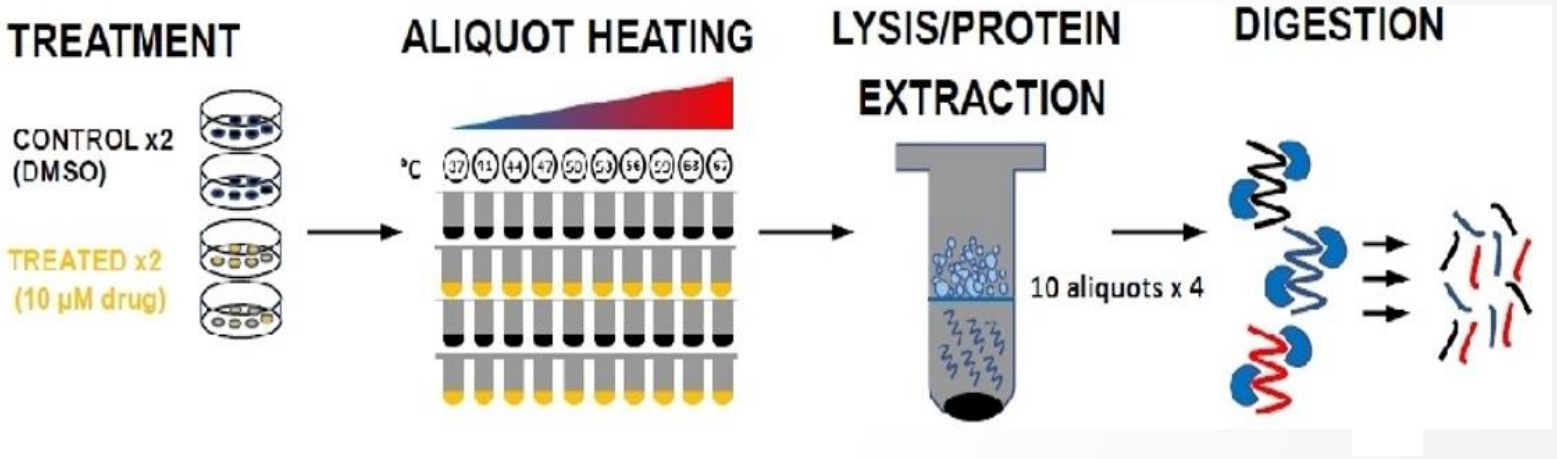
Data files

- Does not contain real data
- To be used just to set up a Study / Analysis
- Do not process the data!



Questions being asked?

- Effect of the drug treatment on protein expression at various temperatures



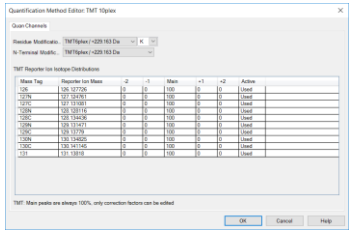
T (°C)	Channel
37	126
41	127N
44	127C
47	128N
50	128C
53	129N
56	129C
59	130N
63	130C
67	131N
Pool	131C

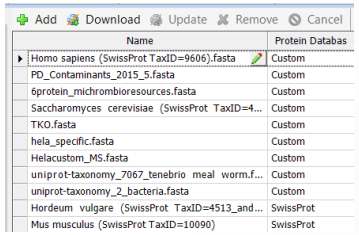
General Quan Experiment Flowchart

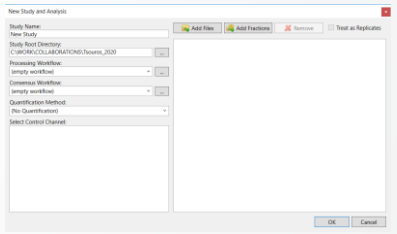
Edit quan method

Add the FASTA db

Create a Study



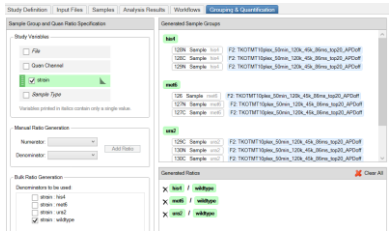


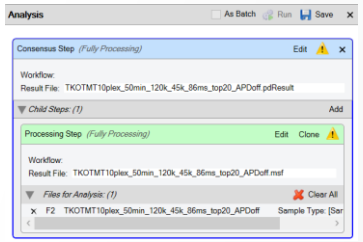


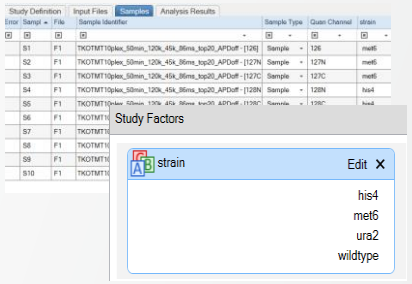
Define Study Factors
Edit Samples table

Edit workflows

Select Study Variables
Add ratio







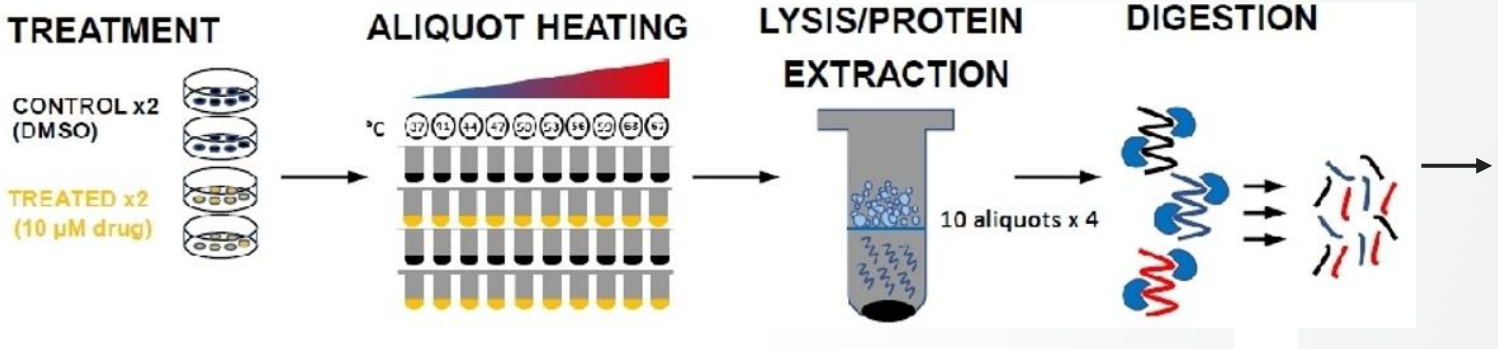
Nested vs Non-nested?

- Are the differences between the two samplings of the same treatment important?

  Should **NOT** be

  ⬇

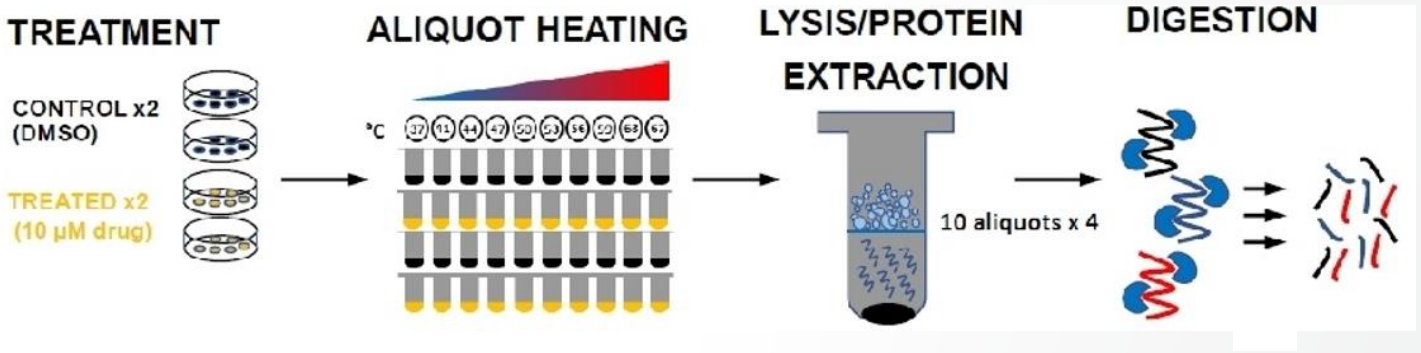
  Non-nested design



T (°C)	Channel
37	126
41	127N
44	127C
47	128N
50	128C
53	129N
56	129C
59	130N
63	130C
67	131N
Pool	131C

Study factors

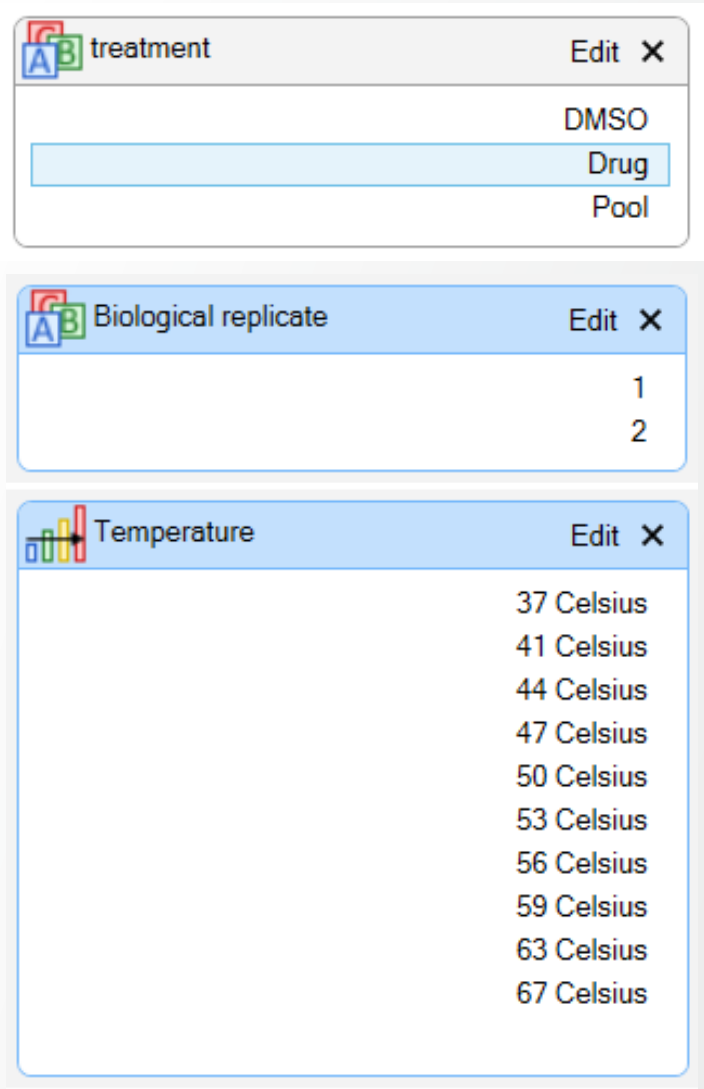
- **Treatment** (DMSO; drug; pool) – *categorical* study factor
- **Biological replicate** (1; 2) – *categorical* study factor
- **Temperature** (37-67; pool) – *numerical* study factor
  - Fractions – use “add fractions”
  - Technical replicate – no dedicated factor needed



T (°C)	Channel
37	126
41	127N
44	127C
47	128N
50	128C
53	129N
56	129C
59	130N
63	130C
67	131N
100	131C

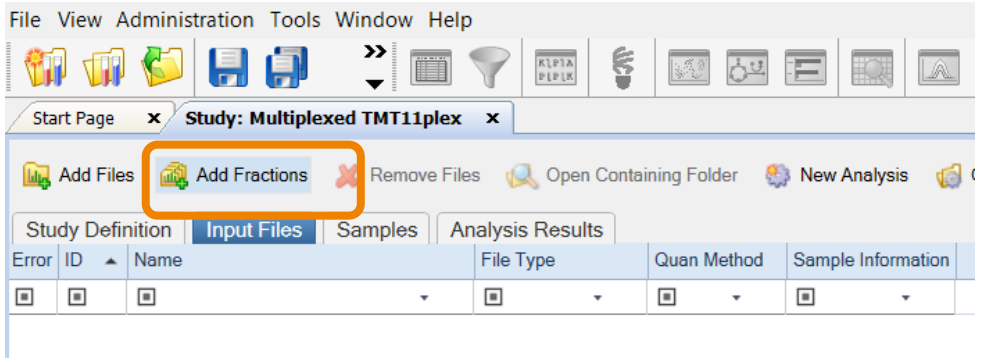
Study factors

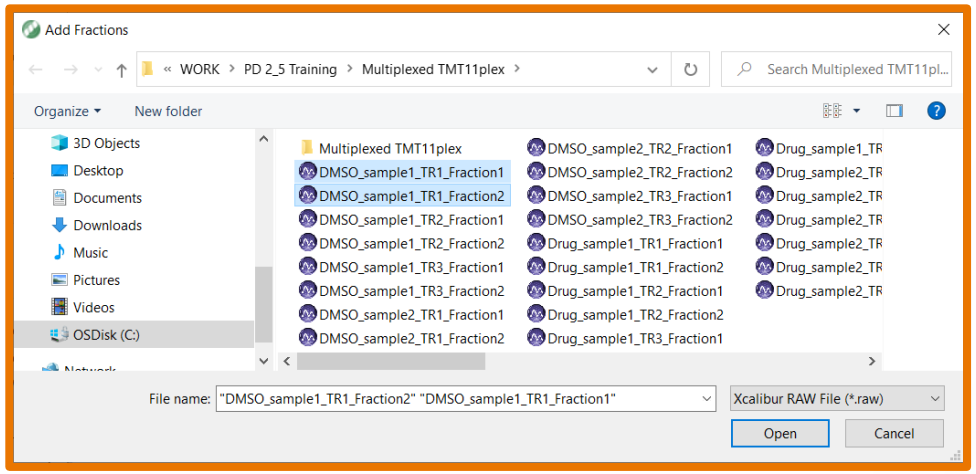
- Treatment (DMSO; drug; pool)
  - *categorical* study factor
- Biological replicate (1; 2)
  - *Categorical* study factor
- Temperature (37-67)
  - *Numerical* study factor
- The order/hierarchy of study factors can be changed later in Grouping&Quantification table

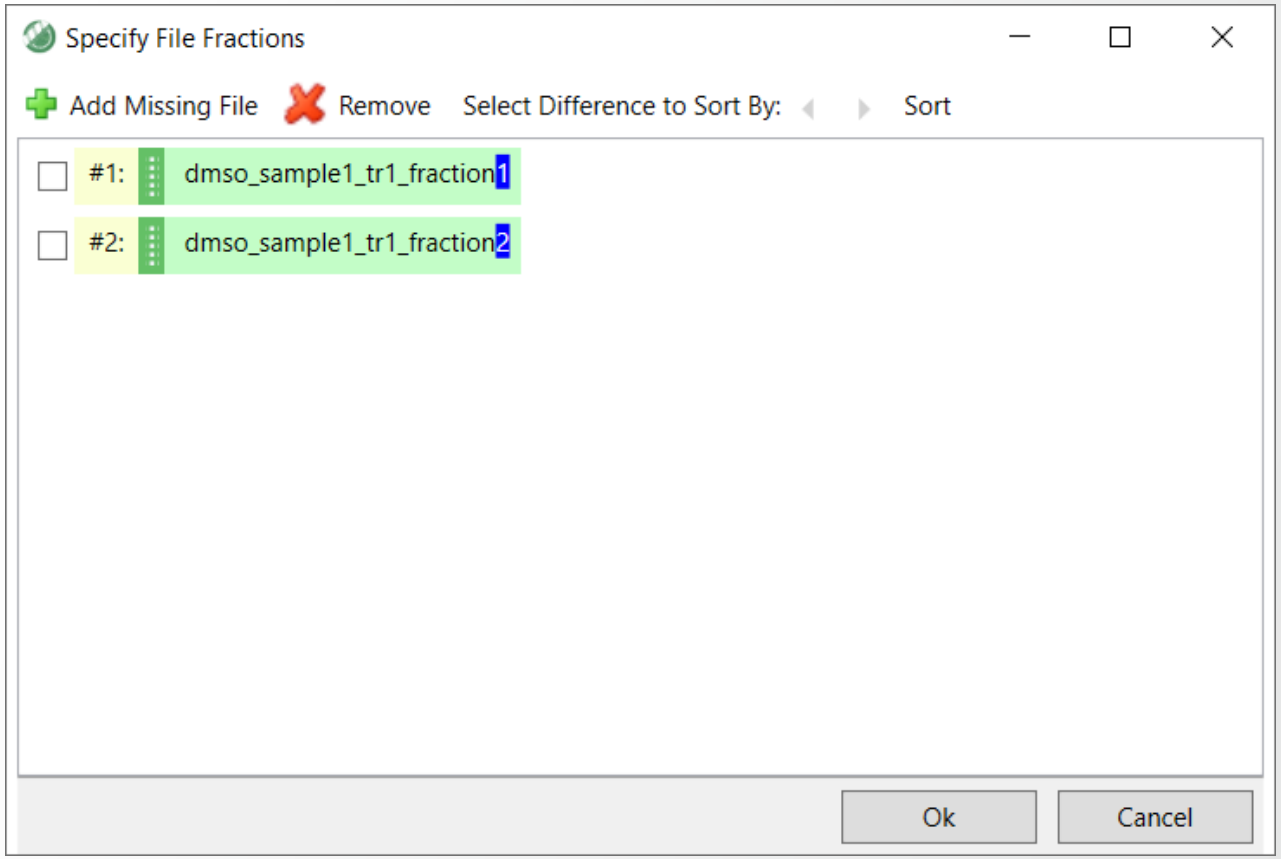


Handling Fractions

- **Input Files** table
- Add Fractions

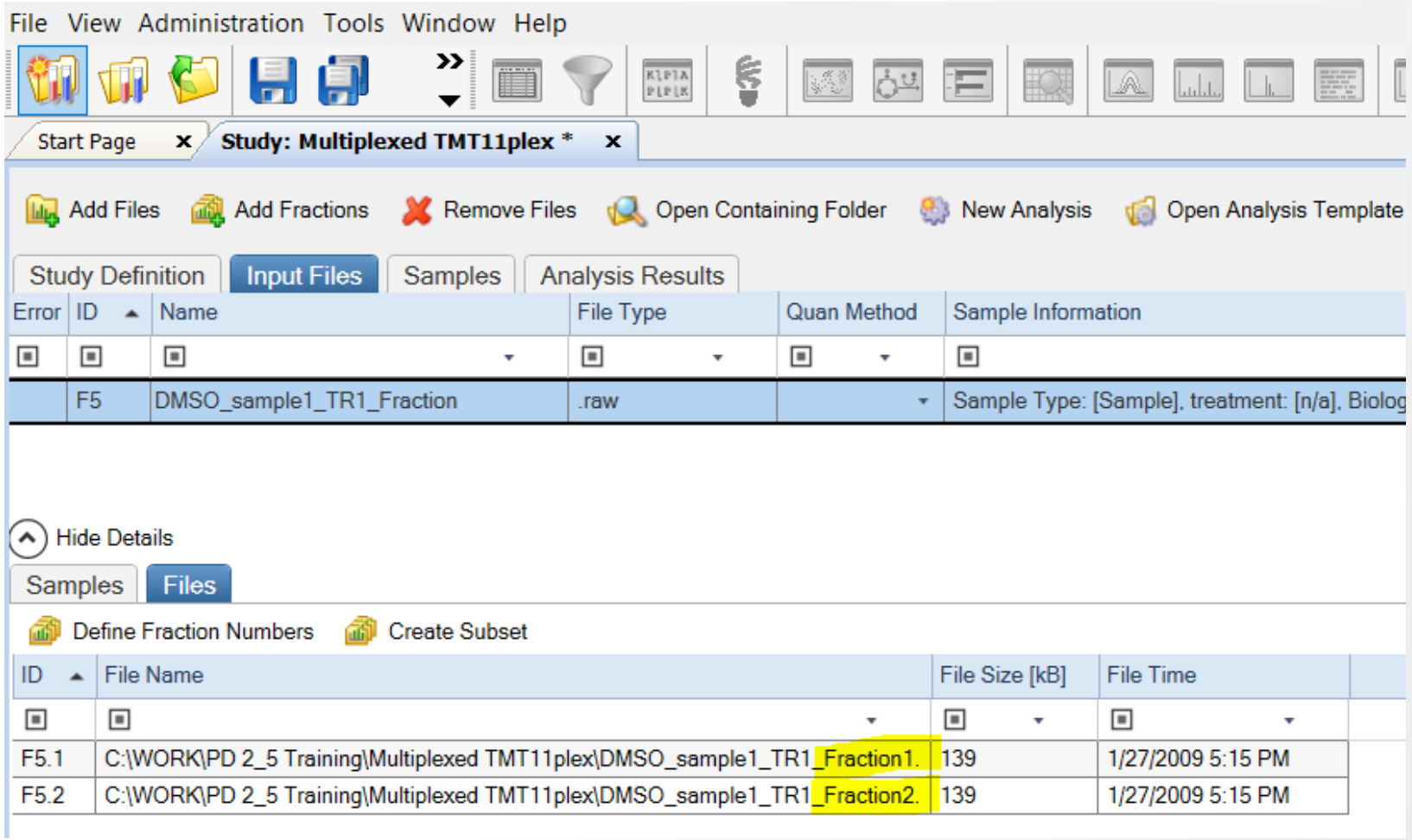






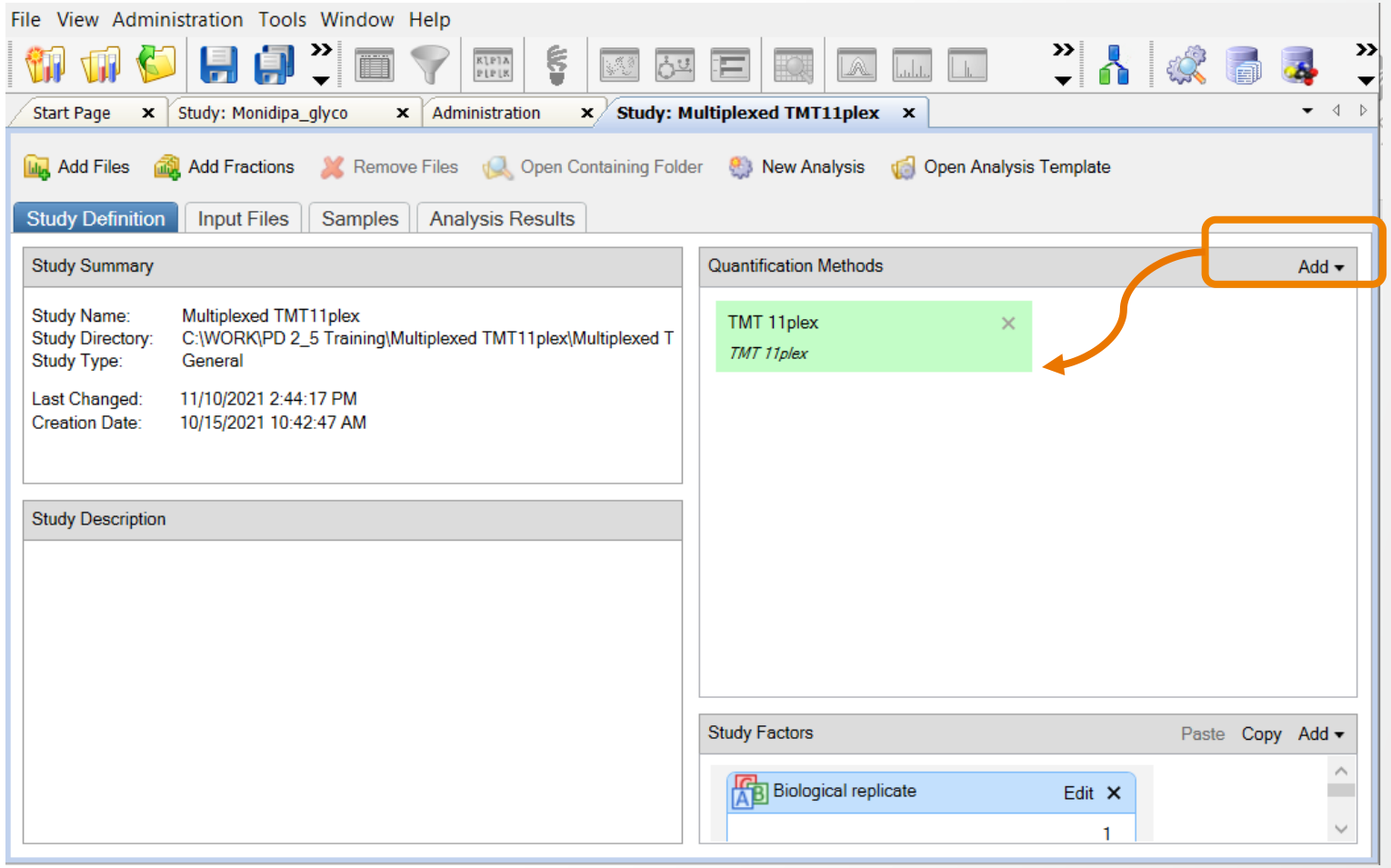
Handling Fractions

- **Input Files** table
- View details of added fractions; then add the rest of the files “as fractions”



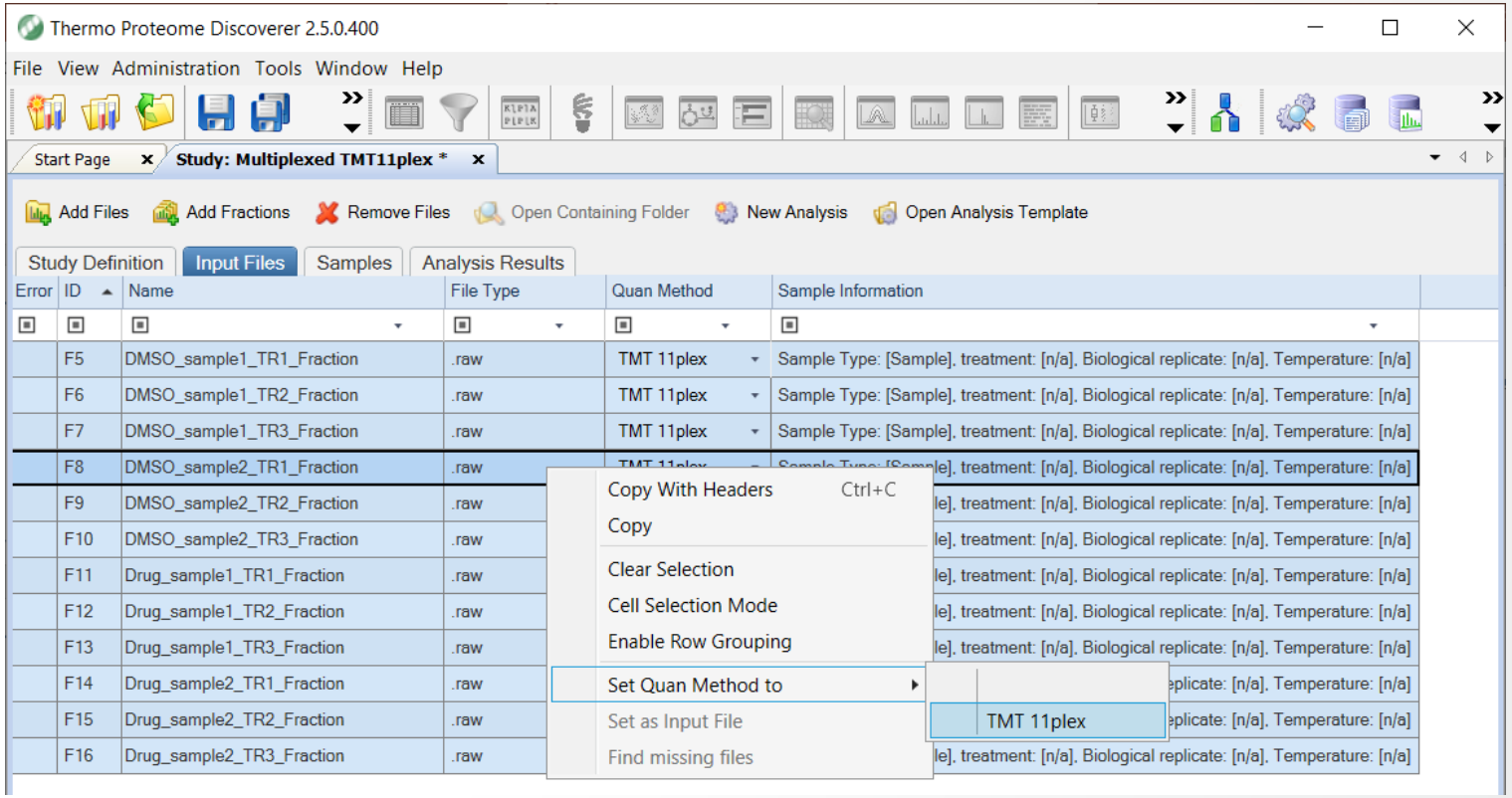
Select quan method

- Using default TMT11plex



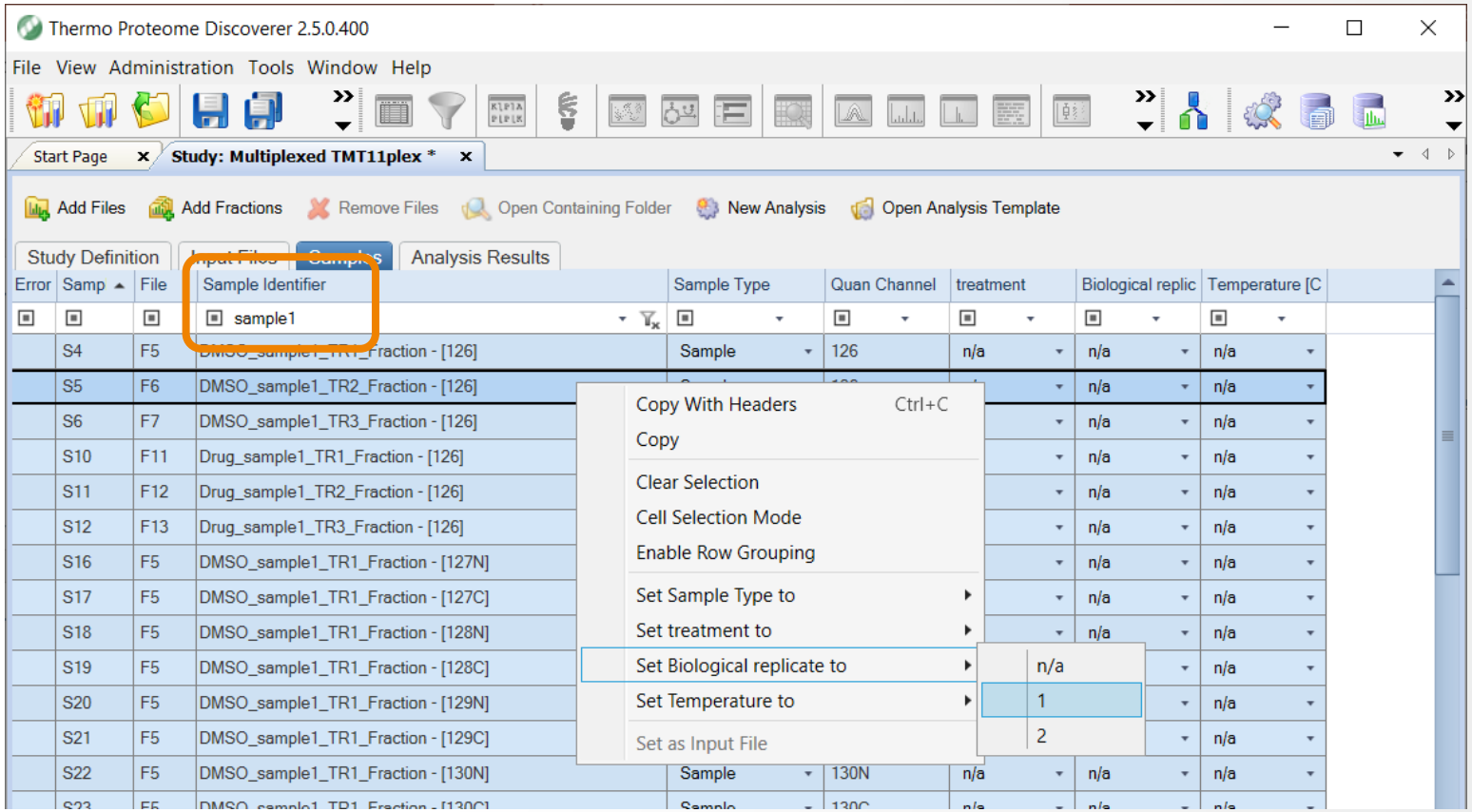
Associate files with quan method

- **Input Files** table
- Highlight all the rows
  - Right mouse click → Set quan method to → TMT11plex



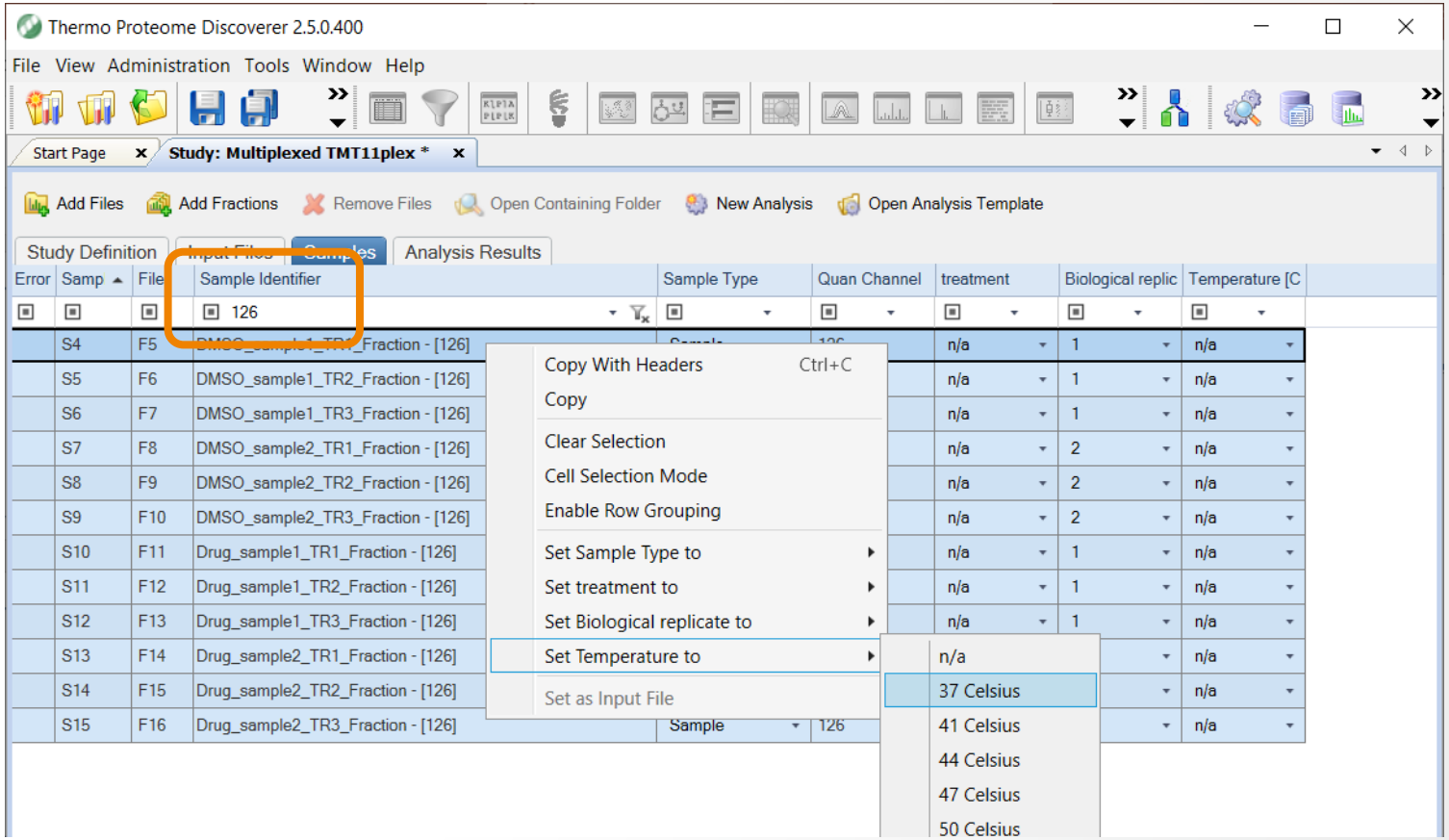
Edit Samples table

- Samples table
- Use sample identifier to select samples of the same group



Edit Samples table

- Samples table
- Use sample identifier to select samples of the same group

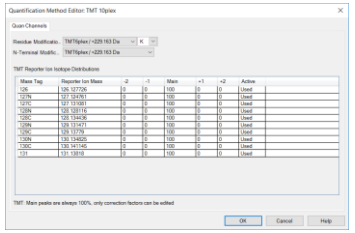


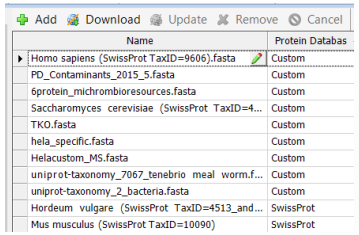
General Quan Experiment Flowchart

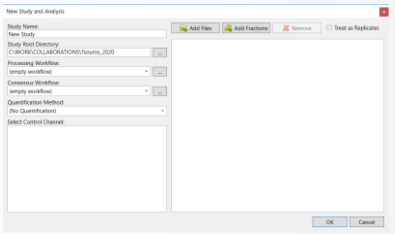
Edit quan method

Add the FASTA db

Create a Study



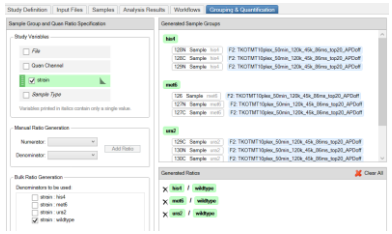


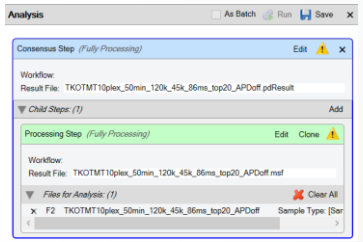


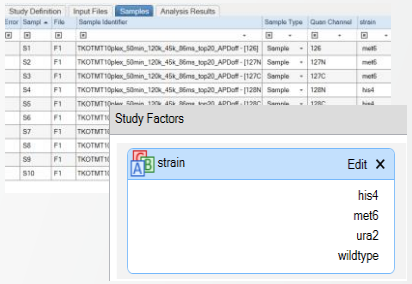
Define Study Factors
Edit Samples table

Edit workflows

Select Study Variables
Add ratio

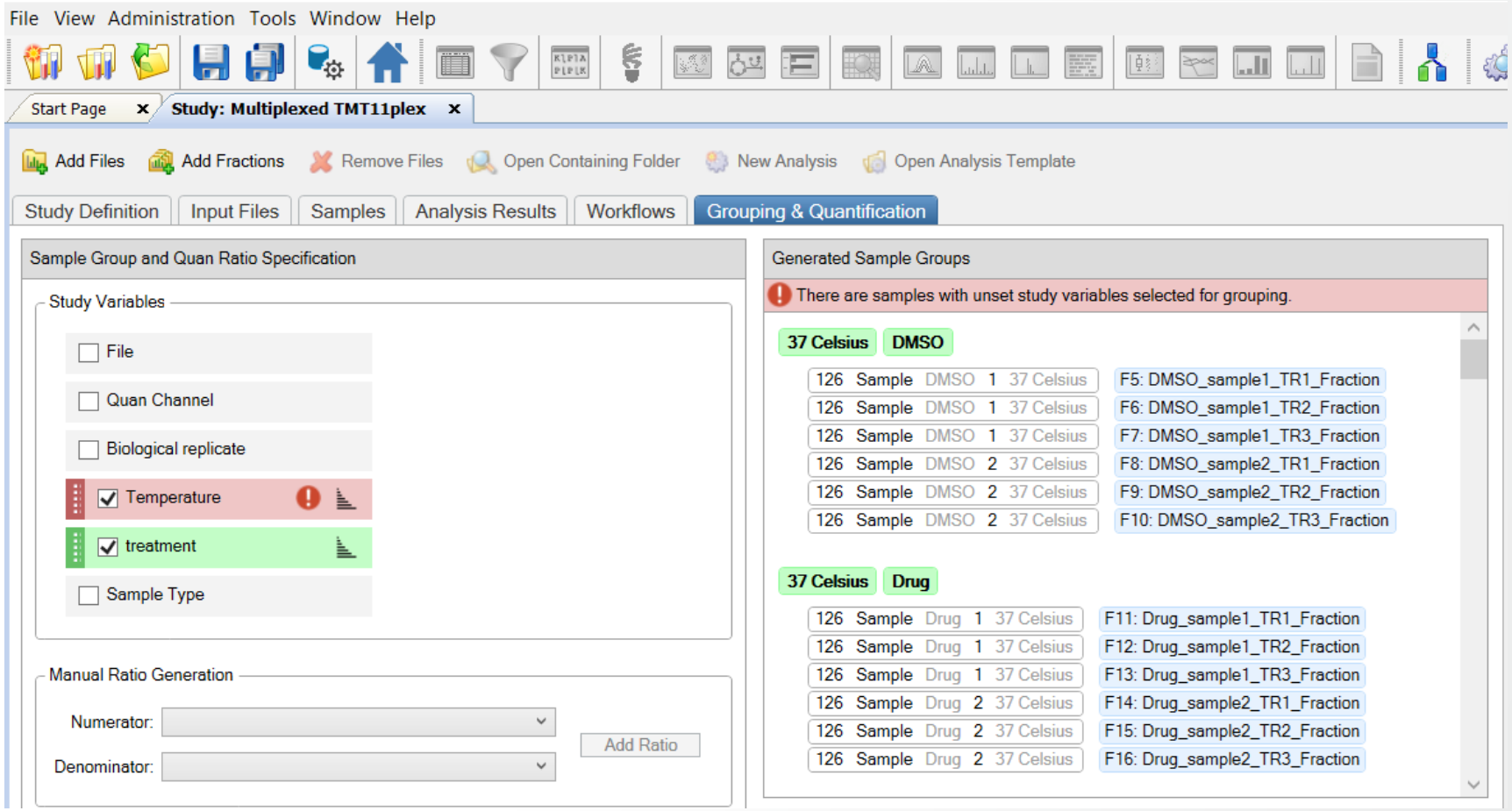






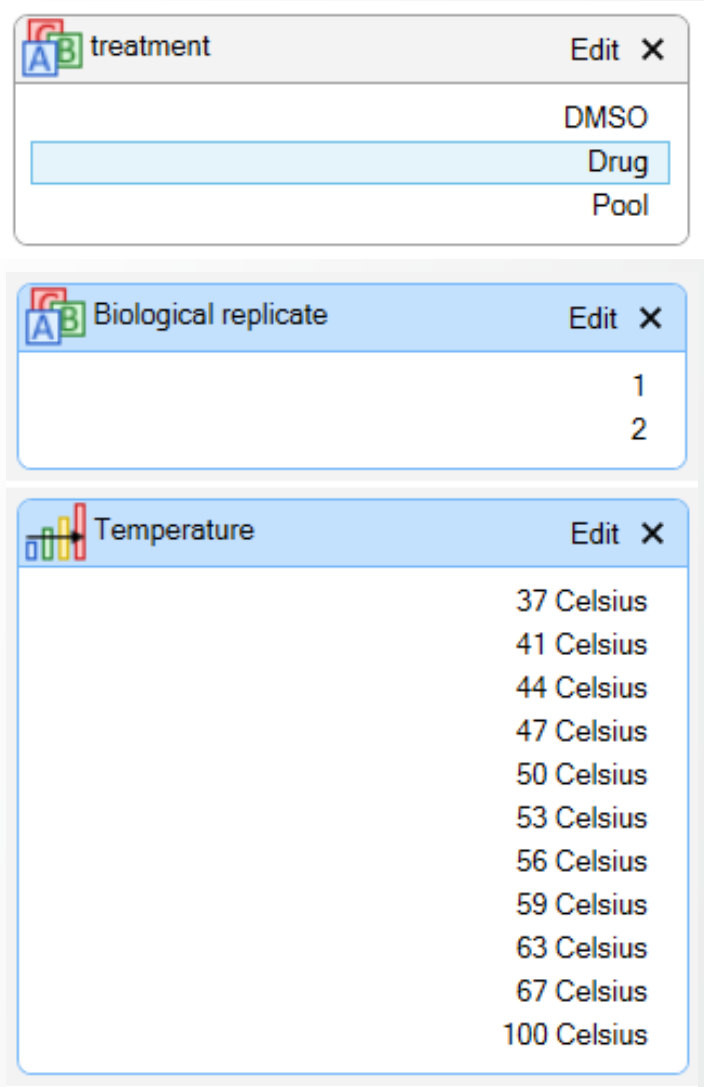
Unset Variables

- PD will not allow you to proceed as some samples have an unassigned variable



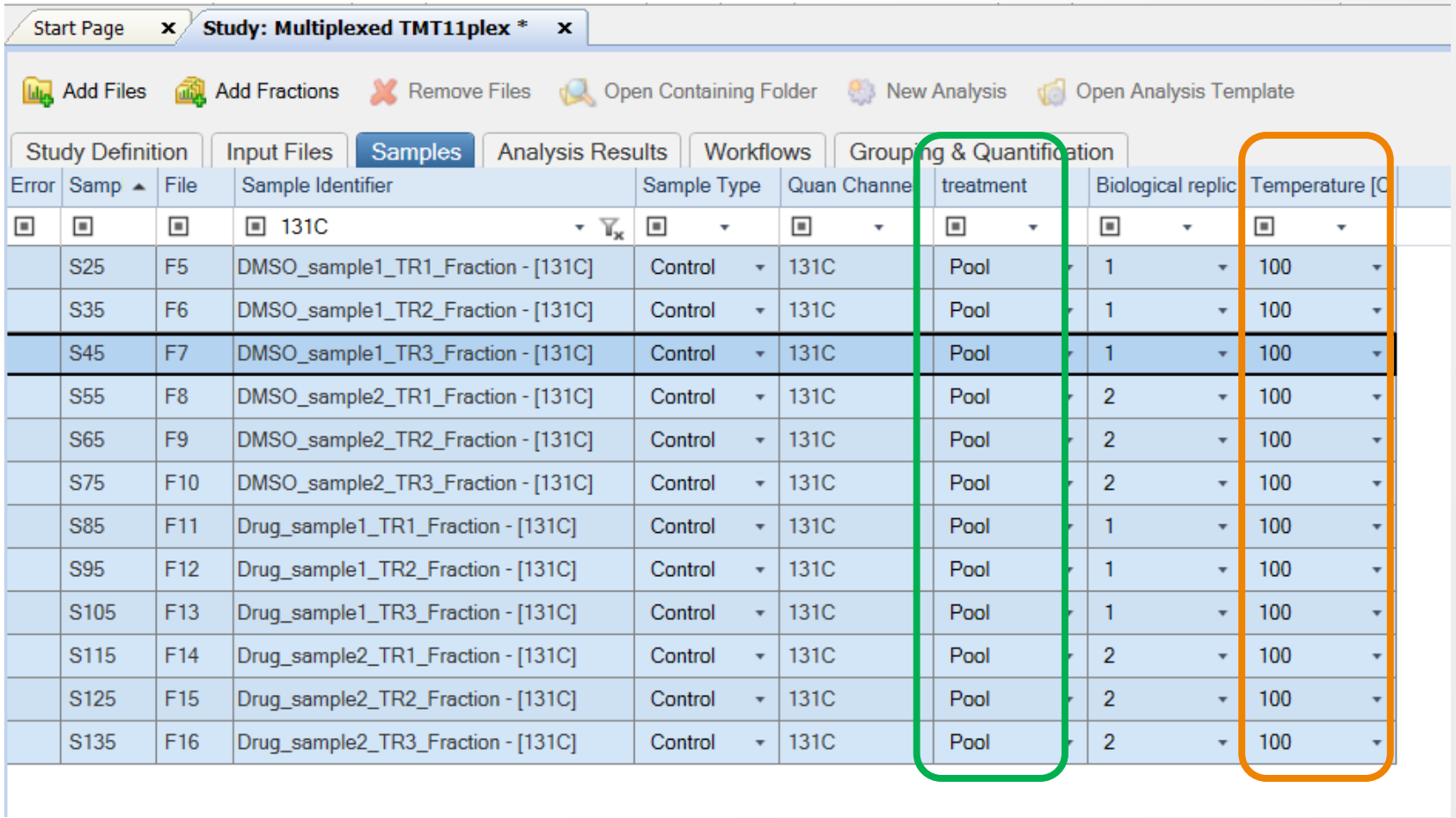
Study factors

- Treatment (DMSO; drug; pool)
  - *categorical* study factor
- Biological replicate (1; 2)
  - *Categorical* study factor
- Temperature (37-67*; value 100 is for "pool" which won't figure in any protein ratio calculation*)
  - *Numerical* study factor
- The order/hierarchy of study factors can be changed later in Grouping&Quantification table



Define Pool sample

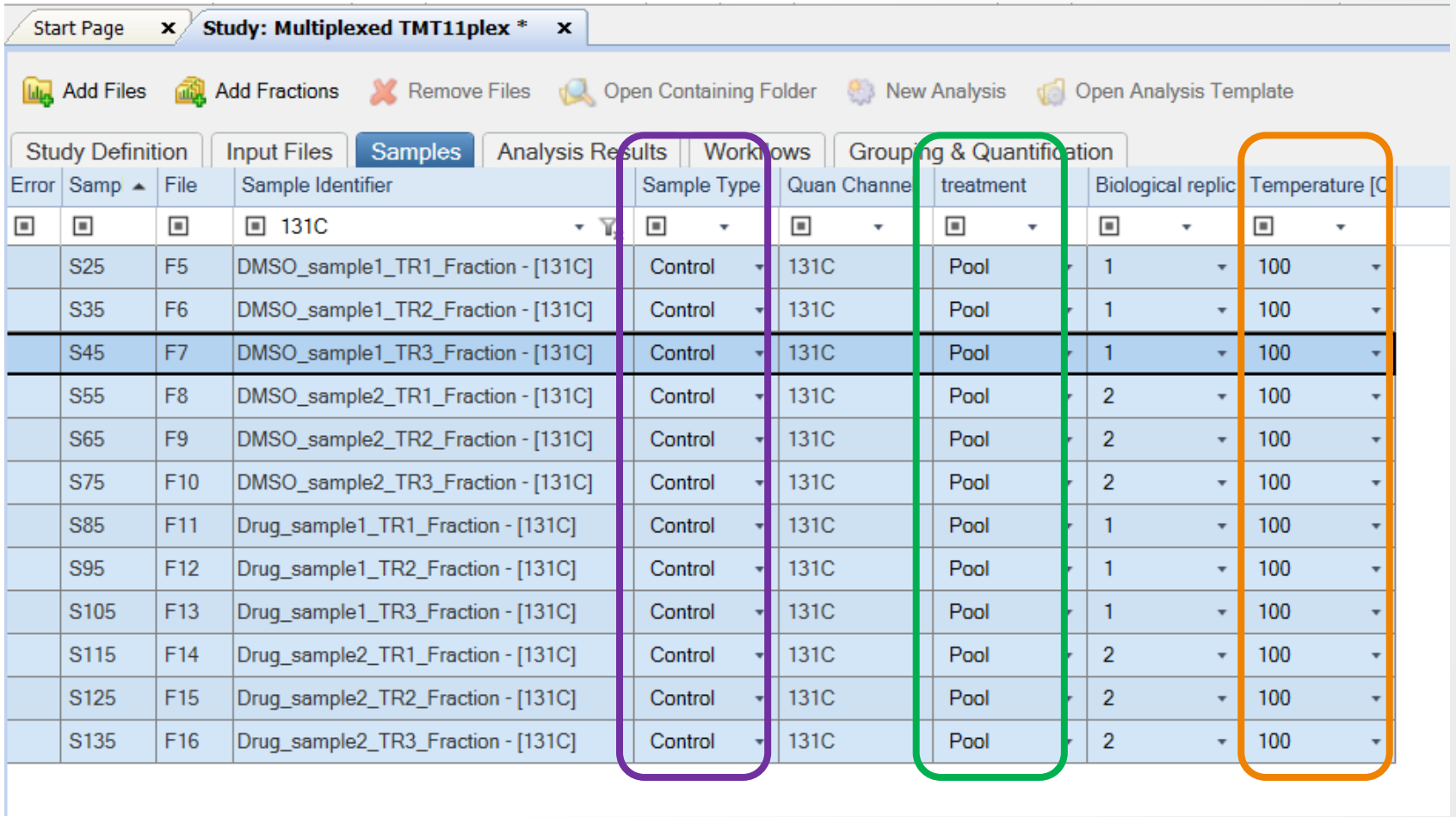
- Set channel 131C as “treatment = pool”; “temperature = 100”



Error	Samp	File	Sample Identifier	Sample Type	Quan Channe	treatment	Biological replic	Temperature [C
			131C					
	S25	F5	DMSO_sample1_TR1_Fraction - [131C]	Control	131C	Pool	1	100
	S35	F6	DMSO_sample1_TR2_Fraction - [131C]	Control	131C	Pool	1	100
	S45	F7	DMSO_sample1_TR3_Fraction - [131C]	Control	131C	Pool	1	100
	S55	F8	DMSO_sample2_TR1_Fraction - [131C]	Control	131C	Pool	2	100
	S65	F9	DMSO_sample2_TR2_Fraction - [131C]	Control	131C	Pool	2	100
	S75	F10	DMSO_sample2_TR3_Fraction - [131C]	Control	131C	Pool	2	100
	S85	F11	Drug_sample1_TR1_Fraction - [131C]	Control	131C	Pool	1	100
	S95	F12	Drug_sample1_TR2_Fraction - [131C]	Control	131C	Pool	1	100
	S105	F13	Drug_sample1_TR3_Fraction - [131C]	Control	131C	Pool	1	100
	S115	F14	Drug_sample2_TR1_Fraction - [131C]	Control	131C	Pool	2	100
	S125	F15	Drug_sample2_TR2_Fraction - [131C]	Control	131C	Pool	2	100
	S135	F16	Drug_sample2_TR3_Fraction - [131C]	Control	131C	Pool	2	100

Define Pool sample

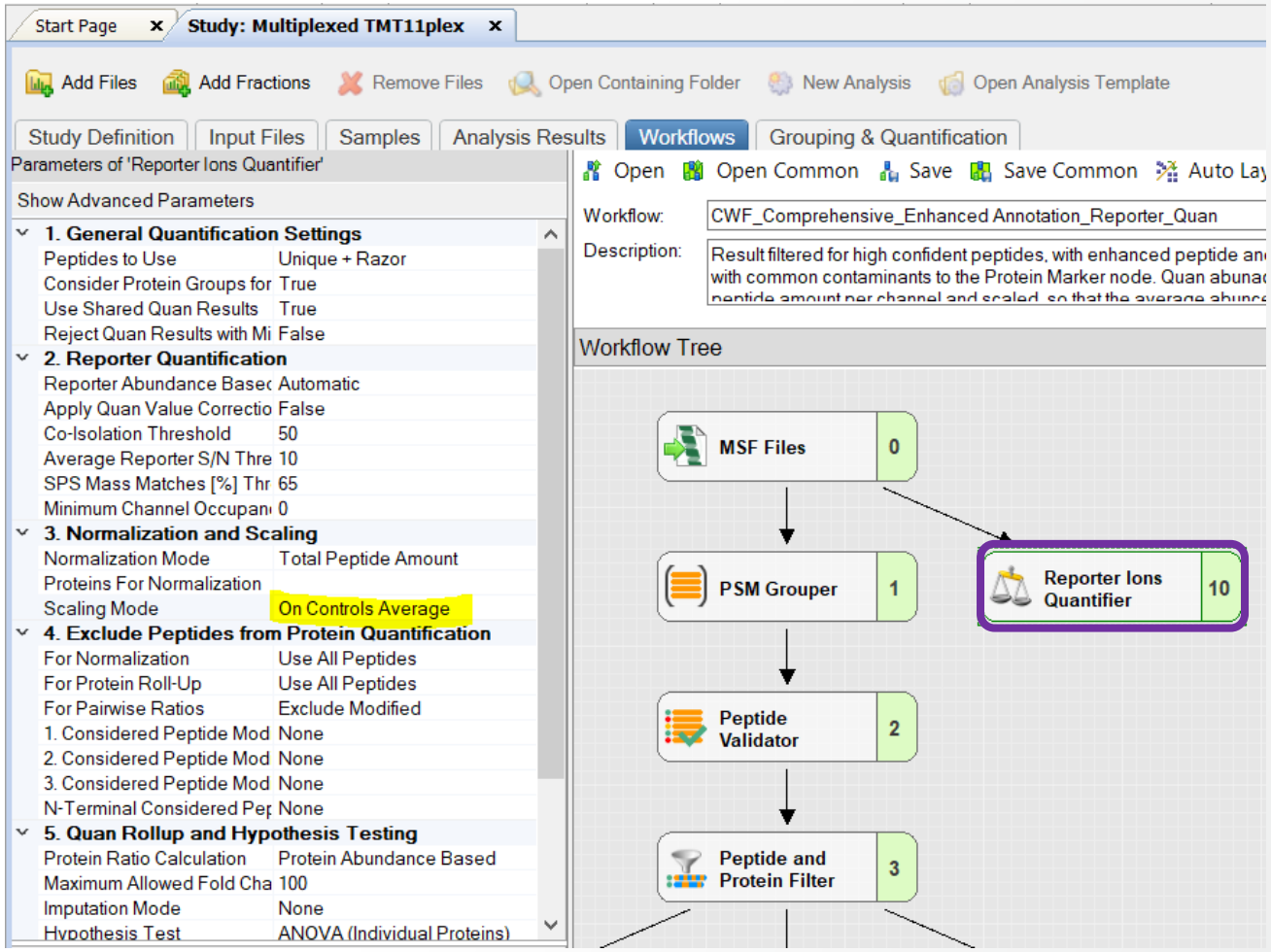
- Set channel 131C as “sample type = control”
  - These samples will be used for normalization across multiple sets



Error	Samp	File	Sample Identifier	Sample Type	Quan Channe	treatment	Biological replic	Temperature [C
			131C					
	S25	F5	DMSO_sample1_TR1_Fraction - [131C]	Control	131C	Pool	1	100
	S35	F6	DMSO_sample1_TR2_Fraction - [131C]	Control	131C	Pool	1	100
	S45	F7	DMSO_sample1_TR3_Fraction - [131C]	Control	131C	Pool	1	100
	S55	F8	DMSO_sample2_TR1_Fraction - [131C]	Control	131C	Pool	2	100
	S65	F9	DMSO_sample2_TR2_Fraction - [131C]	Control	131C	Pool	2	100
	S75	F10	DMSO_sample2_TR3_Fraction - [131C]	Control	131C	Pool	2	100
	S85	F11	Drug_sample1_TR1_Fraction - [131C]	Control	131C	Pool	1	100
	S95	F12	Drug_sample1_TR2_Fraction - [131C]	Control	131C	Pool	1	100
	S105	F13	Drug_sample1_TR3_Fraction - [131C]	Control	131C	Pool	1	100
	S115	F14	Drug_sample2_TR1_Fraction - [131C]	Control	131C	Pool	2	100
	S125	F15	Drug_sample2_TR2_Fraction - [131C]	Control	131C	Pool	2	100
	S135	F16	Drug_sample2_TR3_Fraction - [131C]	Control	131C	Pool	2	100

Reporter Ions Quantifier node

Define scaling “on controls average”

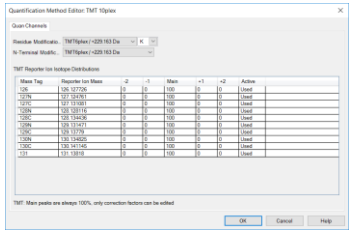


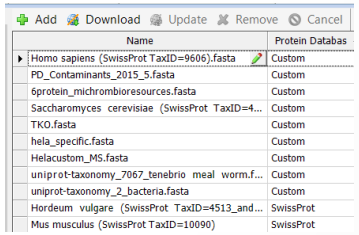
General Quan Experiment Flowchart

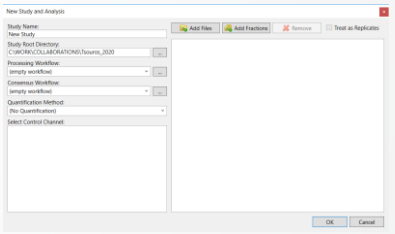
Edit quan method

Add the FASTA db

Create a Study



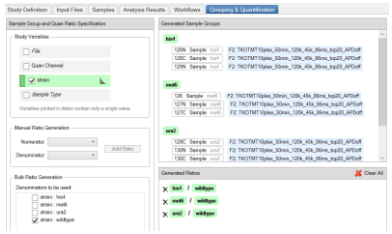


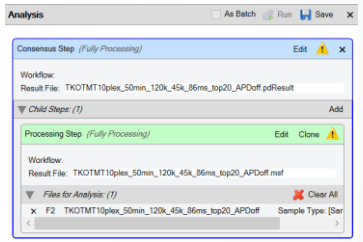


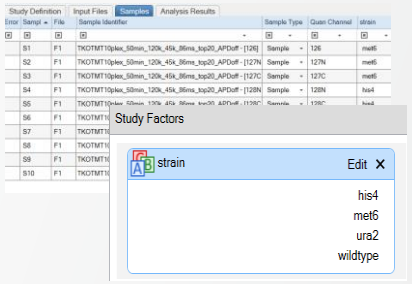
Define Study Factors
Edit Samples table

Edit workflows

Select Study Variables
Add ratio







Q: Is biological variance obscuring the effect of treatment?

Technical replicates presumed to have the lowest variability

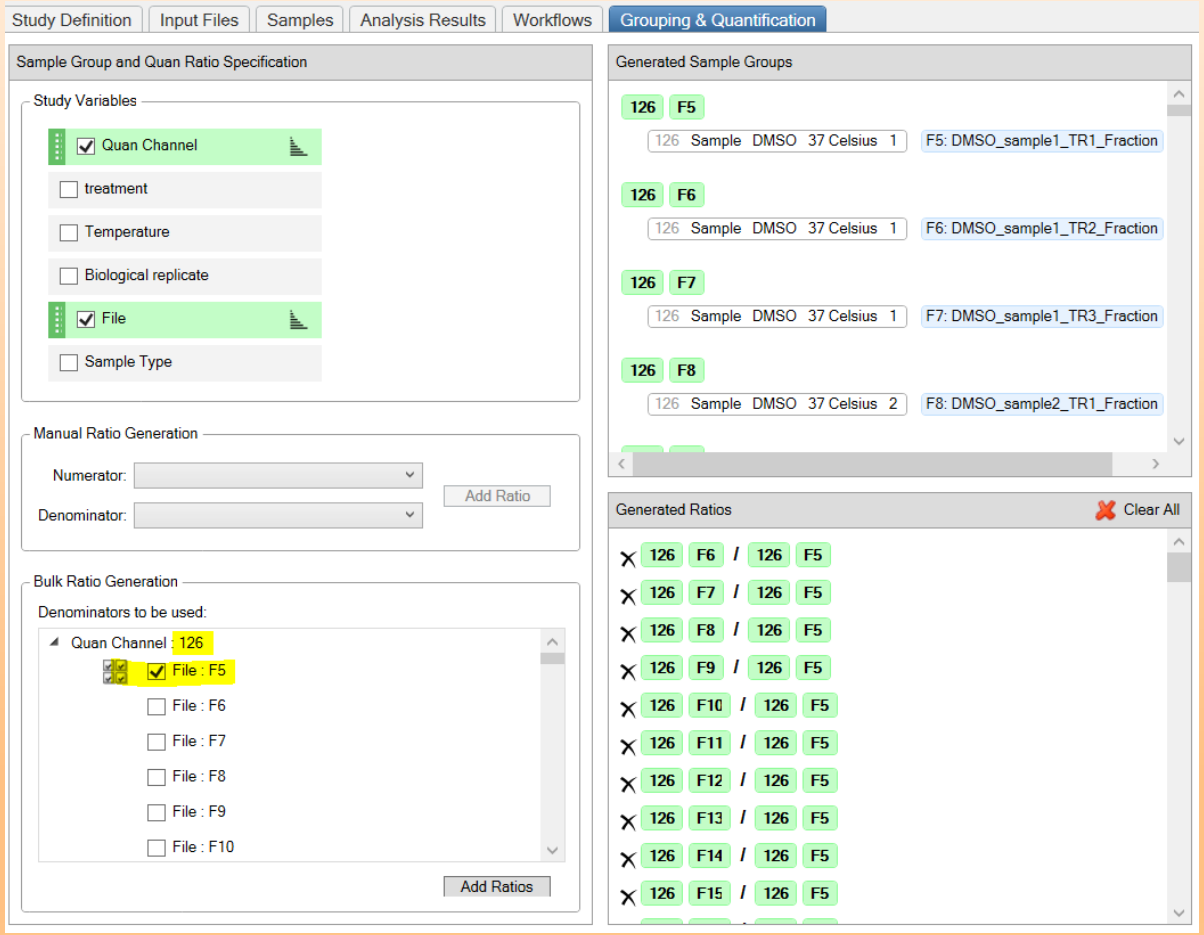
- Set up sample groups containing just a single sample
- Define some ratios (*same sample* as denominator throughout)
- Run the Analysis
- Check the **PCA plot**

- Do samples cluster by channel (=temperature)?

  O.K.

- Or do samples cluster by file (=biological replicate)?

  ?

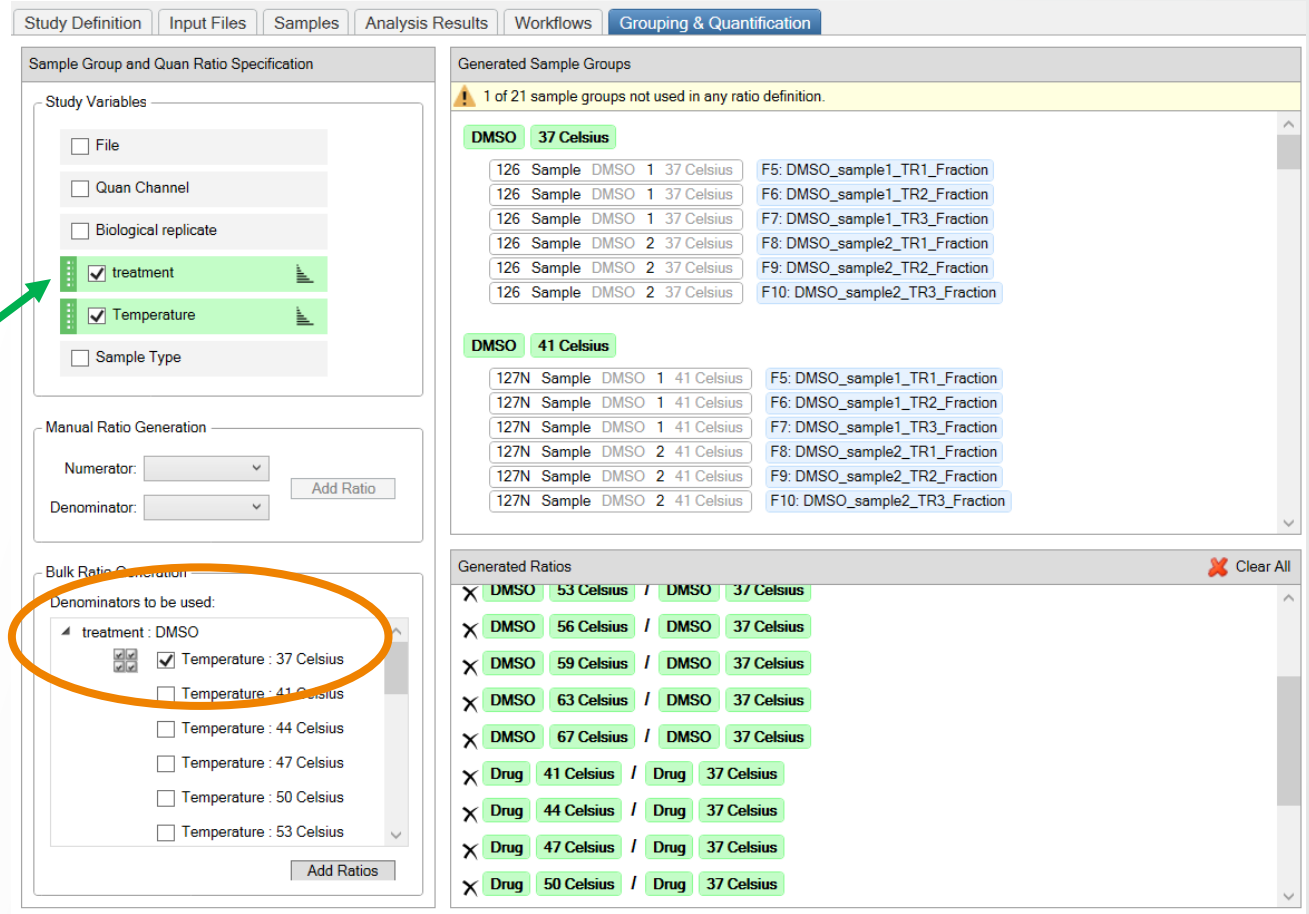


Grouping&Quantification table

Effect of temperature at various treatment conditions

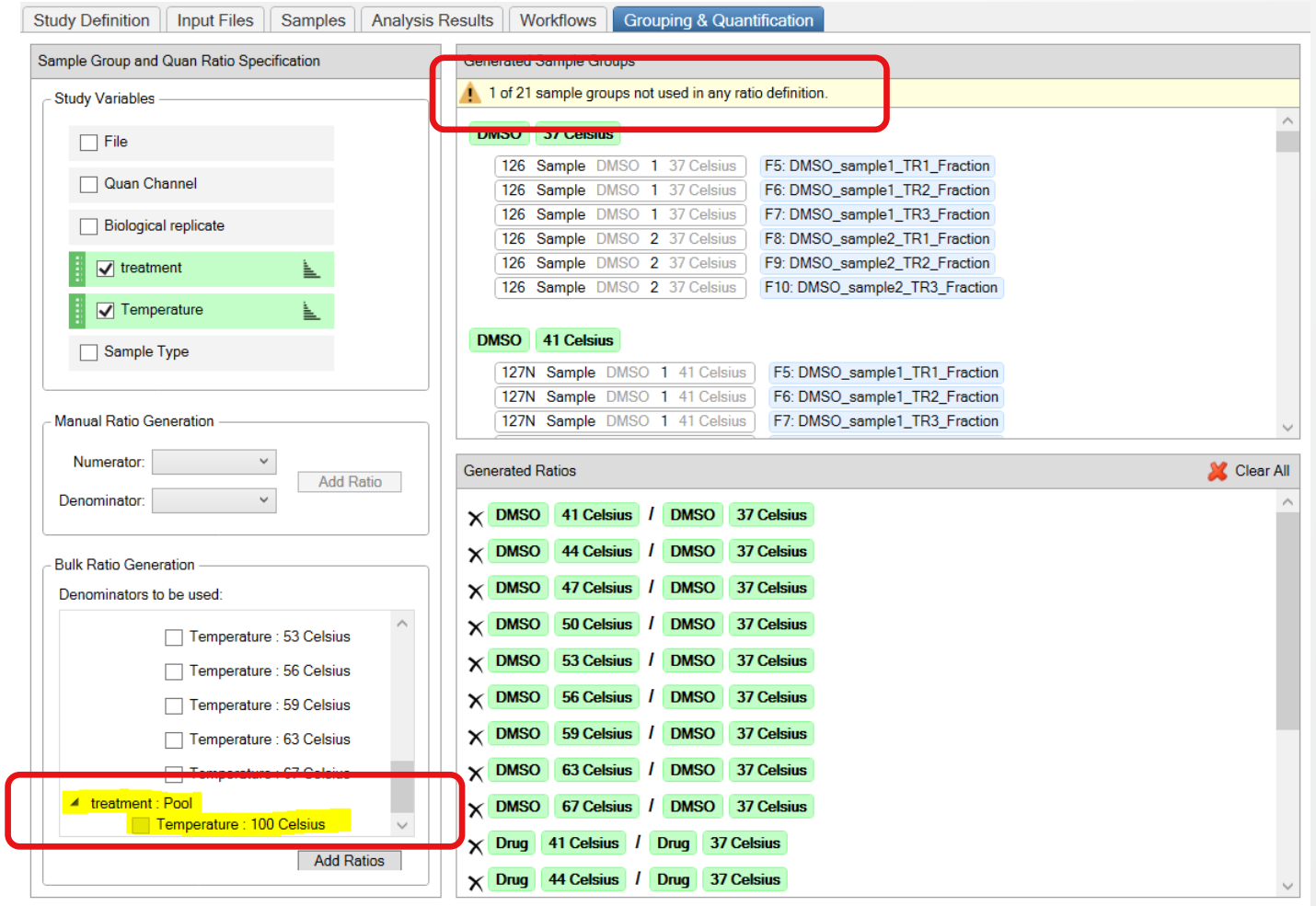
- Primary variable
  “Treatment”
- Secondary variable
  “Temperature”
- Change order of variables if needed

(median 41C DMSO) / (median 37C DMSO)
(median 44C DMSO) / (median 37C DMSO)
(median 47C DMSO) / (median 37C DMSO)
…



Grouping&Quantification table

- One sample group not used in any ratios (the “pool”)



Grouping&Quantification table

Drug treatment effect at various temperatures

- Primary variable

  “Temperature”

- Secondary variable

  “Treatment”

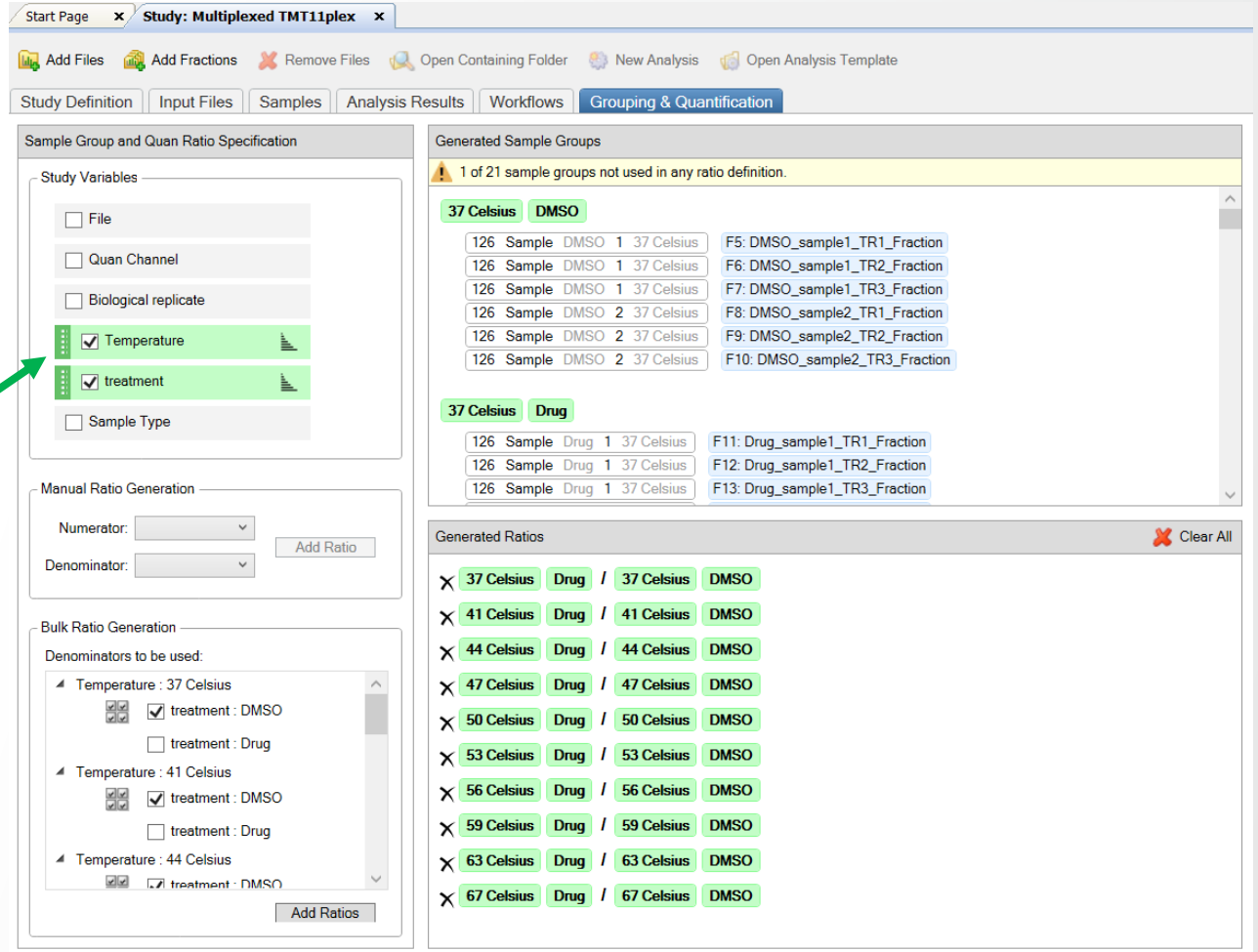
- Change order of variables if needed

(median 37C Drug) / (median 37C DMSO)

(median 41C Drug) / (median 41C DMSO)

(median 44C Drug) / (median 44C DMSO)

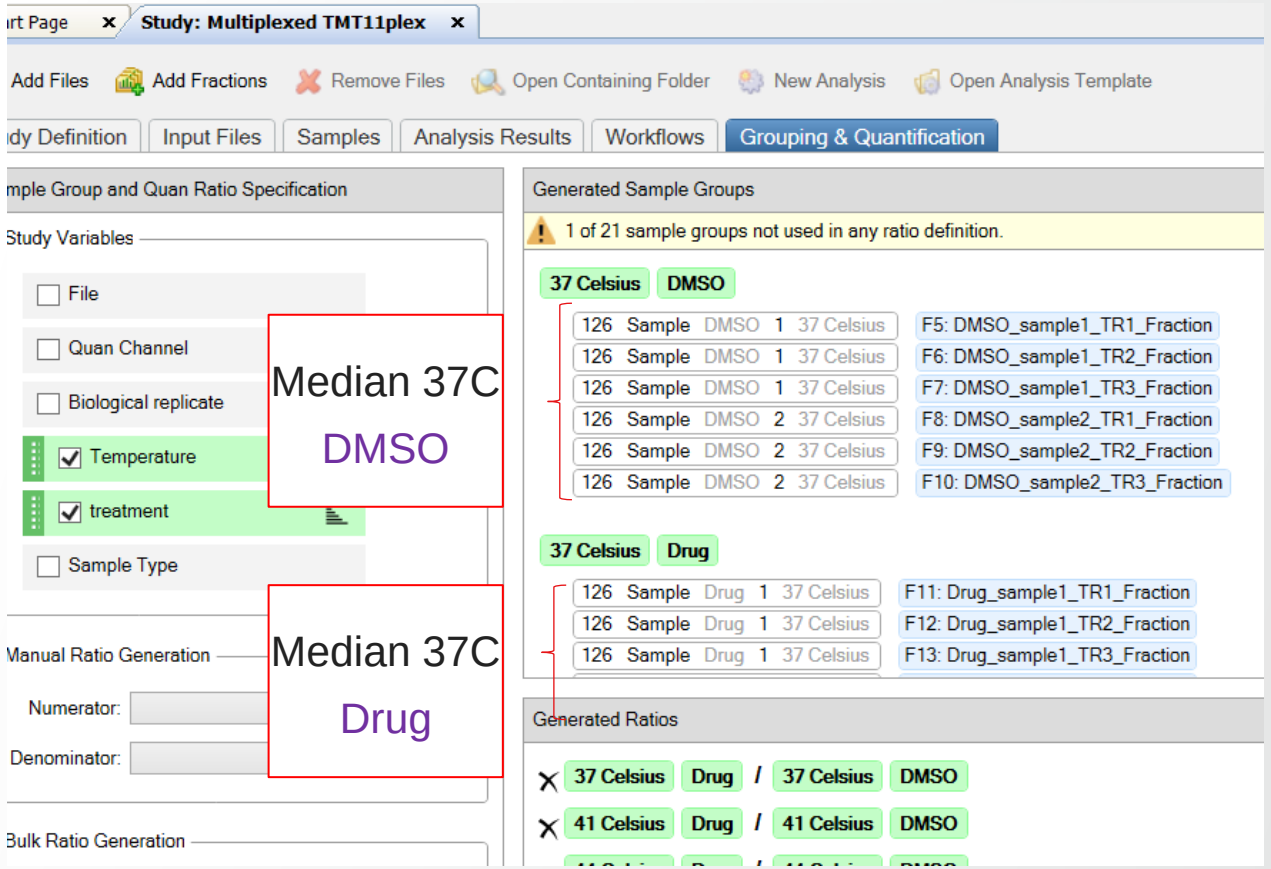
…



Grouping&Quantification table

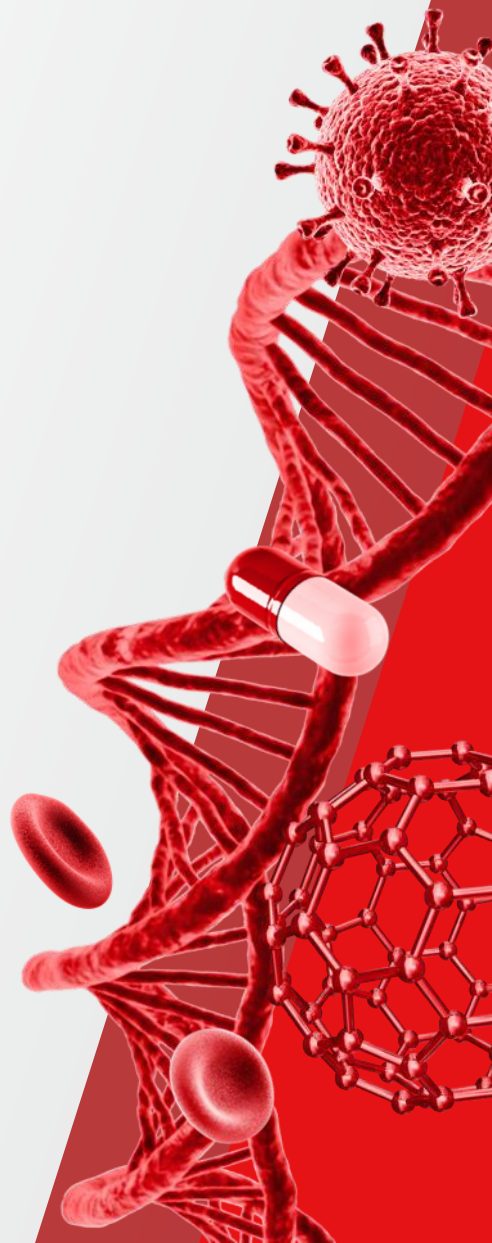
Drug treatment effect at various temperatures

- Protein ratio calculation:
  - PD selects median value representing each sample group
  - In case of even number of values (6), it uses geometrical mean of the two middle values



ThermoFisher SCIENTIFIC

Additional Info

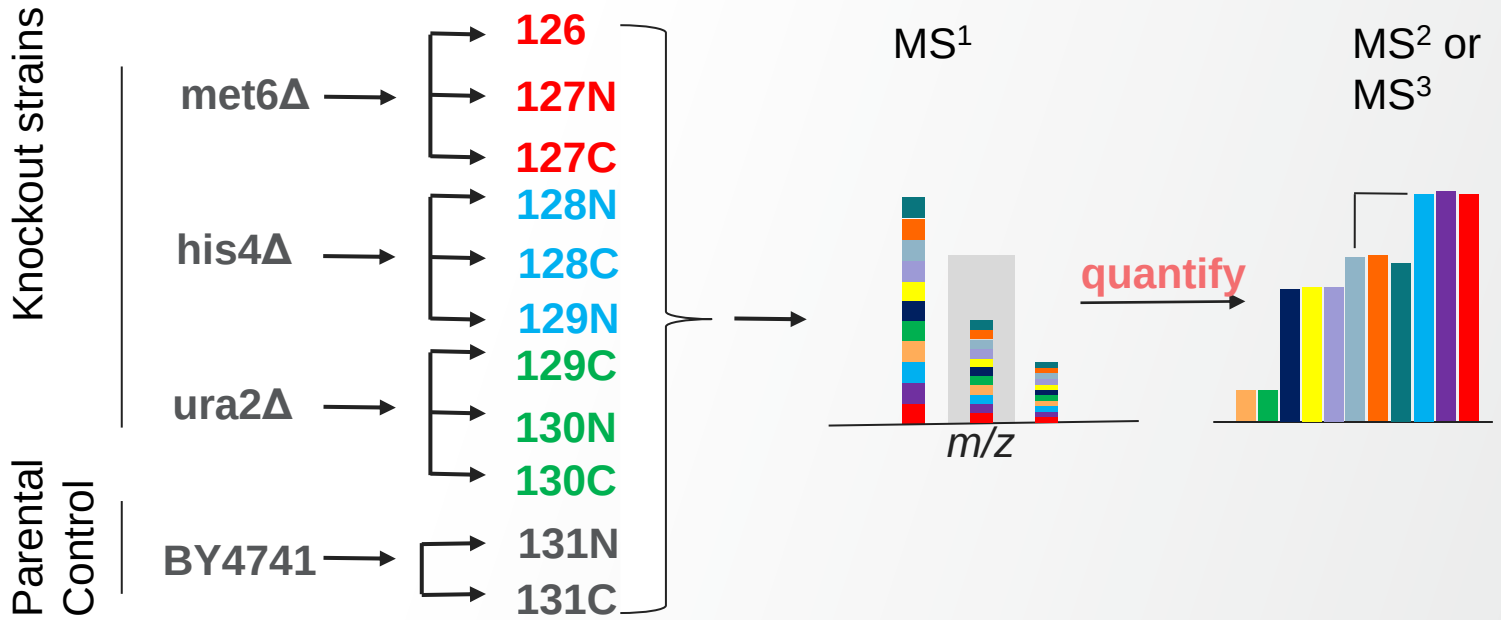


The world leader in serving science

Standard Labeled Sample

For instrument method development
Sample: Thermo Scientific™ Pierce™ TMT11plex Yeast Digest Standard

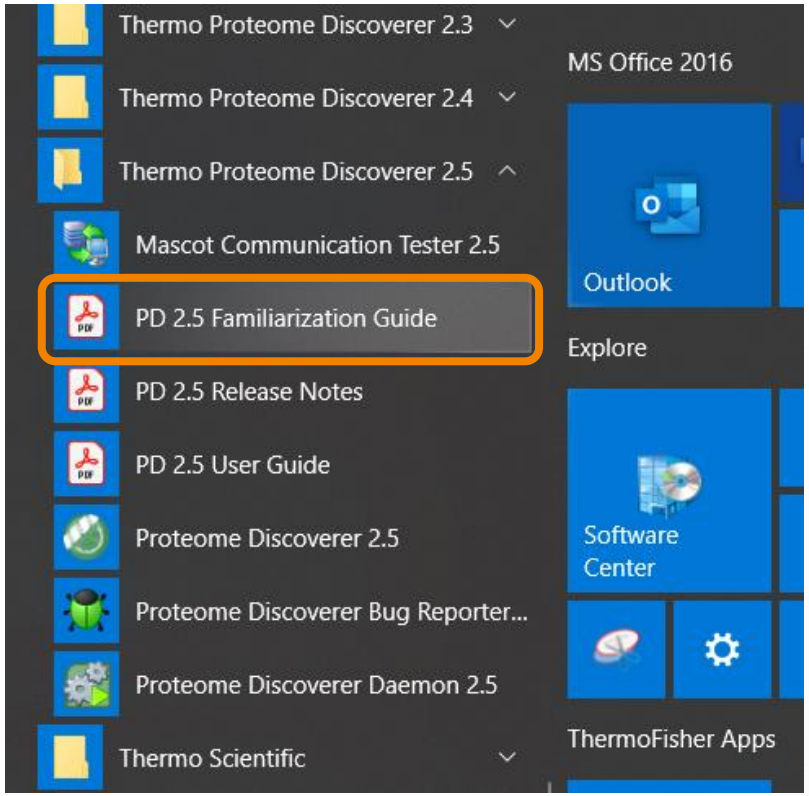
4 strains of *Saccharomyces cerevisiae* → 11 TMT channels technical replicates N=3 or 2 (for BY4741) → Samples combined into 1 LC/MS analysis

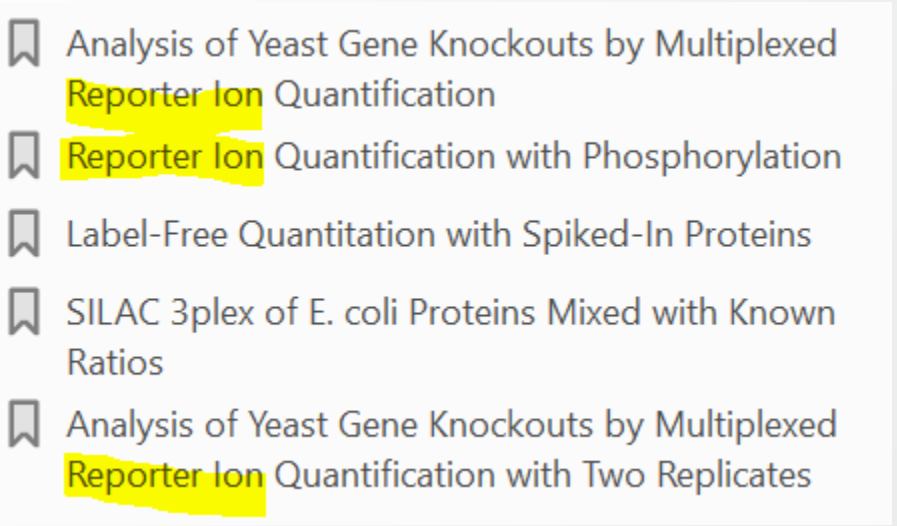


J Am Soc Mass Spectrom. 2016 October ; 27(10): 1620–1625. doi:10.1007/s13361-016-1434-9

More TMT Quan Experiments

- Advanced level experiments in **Familiarization Guide**
- Windows → All Programs → Thermo PD 2.5





PD Support

- pd.support@thermofisher.com

- Provide the following information
  - Problem description
  - Screenshots illustrating the problem, highlighting an example
  - Bug report